

QY 2425 TTAACCTTAAATATTTT 2442

Db	298	TTAATTAAATATTTTT	315
RESULT 2			
ID	T05868		
AC	T05868 standard; DNA; 3399 BP.		
AD	14-AUG-1996 (first entry)		
DE	Chicken leucocytozoan DNA encoding immunogenic protein for vaccines.		
KW	Chicken leucocytozoan; immunogen; recombinant vaccine; protection;		
KW	immunisation; vaccination; ss.		
OS	Chicken leucocytozoan.		
FH	Key	Location/Qualifiers	
FT	cds	1..3399	
FT		/tag= a	
FT	misc_feature	1150..3218	
FT		/tag= b	
FT		/note= "fragment referred to in the claims, for	
FT		use as insert in a recombinant vaccine	
FT		against chicken leucocytozoan disease"	
JN	J07284392-A.		
PN	31-OCT-1995.		
PD	19-APR-1994; 080643.		
PR	19-APR-1994; JP-080643.		
PA	(DOBU-) DOBUTSUYO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.		
PA	(KITA) KITASATO KENKYUSHO SH.		
WP	WPI: 96-006311/01.		
DR	P-PSDB: R97866.		
PT	Chicken leucocytozoan immunogenic protein - used in a recombinant		
PT	vaccine against chicken leucocytozoan disease		
PS	Claim 6; Page 6-9; 35pp; Japanese.		
CC	T05868 encodes a chicken leucocytozoan immunogenic protein, this DNA		
CC	or a fragment of it can be used in a recombinant vaccine to immunise		
CC	against chicken leucocytozoan disease. The DNA is used in a vector		
CC	and operatively linked to an expression regulatory sequence as in		
CC	standard practice.		
SQ	Sequence 3399 BP; 1577 A; 508 C; 798 G; 516 T;		
Query Match 4.0%; Score 97.4; DB 1; Length 3399;			
Best Local Similarity 46.3%; Pred. No. 3.5e-13;			
Matches 320; Conservative 0; Mismatches 371; Indels 0; Gaps 0;			
QY	147	AGGAAAAC TAGGAATCTATGATGCTGTGGTGTGATGGAGATTGATGGATGCCCAA	206
Db	2202	AGAAAAGAAGTAACATCATGAGAGAGAGAGAAAAAGTAACACATGAAGAGAGA	2261
QY	207	AGTTTTATTAGACTTAAGAGAGATCTACTTCAGAGCAGCGTCCGCCCAAGAGGC	266
Db	2262	AGAAAAGTAATACATGAAGAAGAAAAAGAAAGTAATACATGAAGAAGAAAAAGA	2321
QY	267	TGACCACACACTGAGCCCGAGGACGAGTTCCTGTGGAGGCAGAACCCCAGAAATATCGA	326
Db	2322	AGTAACACATGAAGAAGAAAAAGAAAGTAGTAACACATGAAGAAGAAAAAGAAAGTAAC	2381
QY	327	AGATGAAGCAAAAGAAACAATTCAGTCCTCTCTCCATGAATGGTACACGCGAGACATGT	386
Db	2382	ACATGAAGAGAAAAGAGAGTAGTAACATATGAAGAGAAAAAGAAAGTAGTAACACAT	2441
QY	387	TGAGGGAGAGACTTGCAACAGAGAGATGGACCCACAGGAGAACACACAGAGAGATGA	446
Db	2442	AGAAAGAAAAAGTAACACATGAAGAAGAAAAAGAAAGTAGTAACACATGAAGAGAAA	2501
QY	447	TGAGTTTCTTATGGGACTGATGATAGTAGTAGATTTCAGACCCCTGGAACTGAAATATC	506
Db	2502	AGAAAGTAGTAACATGAAGAAGAAAAAGTAACATGAAGAAGAAAAAGTAAC	2561
QY	507	TCATGAAGAACCCAGCAGTAGTTACACGCTGGAAGAGACAGTTTCACAGACTGTAATCA	566
Db	2562	ACATGAAGAGAAAAAGTAACATATGAAGAGAGAGAGAGAGAAAAAGTAAC	2621
QY	567	GGATPATGGAAGAGATGATGCTGTAGCAGGAGAAATCCAGATTCCAGTGAACCACTAGTA	626
Db	2622	ACATGAAGAAGAGAAAAAGTAACACATGAAGAAGAAAAAGTAACACATGAAGAAGA	2681
QY	627	AGATGAAGAGATTGCACCATGATACAGATGATGTAACATACCAAGTCTATCAGAGAAC	686
Db	2682	AGAAAAGTAATACATGAAGAAGAAAAAGAAAGATGAGGAAGAGAGAAAGAGA	2741
QY	687	AGTATATGAACCTCTAGAAAATGAAGGATAGAAAATCACAGAAAGTAACTCTCCCC	746
Db	2742	AGCAAGAAGAGAGGAAGAAGAAGAAGATGAGGAAGAGAGAGAGAGAGAGAGA	2801
QY	747	GGATAATCTCTGTAGAAGATTCACAGCTAATTTGTAGAAGAGTAGCATTTTCTCTGT	806
Db	2802	AGATGAGGAAGAGGAAGAAGAAGAAAATGAGGAAGAGAGAGAGAGAGAGAGA	2861
QY	807	AGAACAGCAGGAGAGTACCACCAGATACTTAA	837
Db	2862	AGAGAAGAGAGAGAAAAAGAGAGCATGAA	2892
RESULT 3			
ID	Q87587		
AC	Q87587 standard; DNA; 1686 BP.		
DT	19-DEC-1995		
DE	DNA encoding Leucocytozoan protozoa structural protein epitope.		
KW	Leucocytozoan protozoa; structural protein; epitope; vaccine; fowl;		
KW	leucocytozoanosis; treatment; ss.		
OS	Leucocytozoan protozoa sp.		
PN	J07089995-A.		
PD	04-APR-1995.		
PF	10-SEP-1993; 226078.		
PR	10-SEP-1993; JP-226078.		
PA	(DOB-) DOBUTSUYO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.		
PA	(NISS-) NISSEIKEN KK.		
DR	WPI: 95-167252/22.		
DR	P-PSDB: R70491.		
PT	Immune inducing polypeptide against Leucocytozoan protozoa - useful		
PT	in production of vaccines for treatment of leucocytozoanosis in		
PT	fowl.		
PS	Claim 1; Page 12-14; 20pp; Japanese.		
CC	Q87587-89 encode polypeptides having a whole or partial epitope of a		
CC	structural protein of Leucocytozoan protozoa (see R70491-93). The		
CC	polypeptides and DNA encoding them are useful in the production of		
CC	vaccines for the treatment of leucocytozoanosis of fowl.		
SQ	Sequence 1686 BP; 915 A; 129 C; 382 G; 260 T;		
Query Match 3.9%; Score 95.2; DB 1; Length 1686;			
Best Local Similarity 46.9%; Pred. No. 8.4e-13;			
Matches 335; Conservative 0; Mismatches 373; Indels 6; Gaps			
QY	137	AGGAAGTCTTAGGAAAAC TAGGAATCTATGATGCTGTGATGGATGGAGATTTGTATG	196
Db	125	AAGAAGAACAAAGAAGAAGAACCAAGAACAAAGAACAAAGAAATCGTAGAAG	184
QY	197	ATGATGCCAAACTTTTATTAGGACTTAAAGAGAGATCTACTTCAGAGCCAGCAGTCCC	256
Db	185	ACAAGAACAAGATGAAGAAGAACCAAGAAAGAGAGATGAAGAAGAGAGAGAGAAA	244
QY	257	CAGAAGAGGCTGAGCCACACACTGAGCCCGAGGAGCAGGTTCCTGTGGAGGAGAAC	316
Db	245	AGAAGAAGAACAG	304
QY	317	AGAAATCGAAGATGAAGCA	

[illegible]

Qy	412	GATGGACCCACAGAGAACACAAACAAGAGAGATGATGAGTTTCTTATGGCGACTGATGTA	471
Db	3821	GTTGAAGAAGTAGAAGAAAATGTAGAAGAAAATGTAGAAGAAAATGTGTAGAGAAAATGTT	3880
Qy	472	GATGATAGATTTGAGACCCCTGGAACTGAAGTATCTCATGAAGAAACCGACCATAGTTAC	531
Db	3881	GAAGAAGTAGAAGAAAATGTAGAAGAAAATGTAGAAGAAAATGTAGAAGAAAATGTTGAA	3940
Qy	532	CACGTGAAGAGACAGTTTCACAAGACTCTAATCAGGATATGGAAGAGATGATCTCTGAG	591
Db	3941	GAATACTTGAAGAAAATGTTGAAGAAAATGTAGAAGAAAATGTAGAAGAAAATGTTGAA	4000
Qy	592	CAGGAAAATCCAGATTCCAGTGAACCAAGTAGTAGAAGATGAAGATTCACCATGATACA	651
Db	4001	GAATATGATCAAGAAAATGTTGAAGAAGTAGAAGAAAATGTAGAAGAAAATGTAGAGAA	4060
Qy	652	GATGATGTACATACCAAGTCTATGAGGAACAAGCAGATATATGAACCTCTGTAGAAAATGAA	711
Db	4061	AATGTTGAAGAAAATGTAGAAGAAAATGTTGAAGAAGTAGAAGAAAATGTAGAAGAAAAT	4120
Qy	712	GGGATAGAAATCACAGAAGTAACTGCTCCCCC-----TGAGGATAATCCTGTAGAAGAT	765
Db	4121	GTAGAAGAAAATGTAGAAGAGATGTTGAAGAGAAATGTTGAAGAGAAATGTTGAAGAAATAT	4180
Qy	766	TCACAGGTAATTTAGAAGAAGTAAAGCATTTTCTCTGGAAGAACACGACAGAAAGTACCA	825
Db	4181	GATGAAGAAAATGTTGAAGAACACAAATGAAGAAATATGATGAATAAAAAAAATATATA	4240
Qy	826	CCAGATCTTAAAGCTTCAAAA	847
Db	4241	TTAAAGTTTTAAATTTTTTATAAA	4262

RESULT 6

Q87588 Q87588 standard; DNA; 996 BP.

ID Q87588: AC Q87588: 19-DEC-1995 (first entry)

DT DE DNA encoding Leucocytozoan protozoa structural protein epitope.

DE CC leucocytozoan protozoa; structural protein; epitope; vaccine; fowl;

KW leucocytozoan protozoa; treatment; ss.

KW OS Leucocytozoan protozoa sp.

OS PN J07089995-A.

PN PD 04-APR-1995.

PD PF 10-SEP-1993; 226078.

PF PR 10-SEP-1993; JP-226078.

PR PA (DOB-) DOBUTSUO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.

PA PA (NISS-) NISSEIKEN KK.

PA DR WPI: 95-167252/22.

DR PT Immune inducing polypeptide against Leucocytozoan protozoa - useful

PT PT in production of vaccines for treatment of leucocytozoanosis in

PT fw1.

fw1 Claim 1; Page 14-15; 20pp; Japanese.

PS CC Q87587-89 encode polypeptides having a whole or partial epitope of a

CC CC structural protein of Leucocytozoan protozoa (see R70491-93). The

CC CC polypeptides and DNA encoding them are useful in the production of

CC CC vaccines for the treatment of leucocytozoanosis of fowl.

CC SQ Sequence 996 BP; 500 A; 94 C; 221 G; 181 T.

SQ

Query Match 2.8%; Score 68.6; DB 1; Length 996;
Best Local Similarity 50.9%; Pred. No. 7.8e-07;
Matches 254; Conservative 0; Mismatches 224; Indels 21; Gaps 3;

[illegible]

QY	412	--GATGGACCCACAGGAGAACCCACCAAGAGGATGAATGAGTTCTTATATGGCGACTGATG	469
Db	530	TGGAGAGAGATGTAGTAGAAGAACTTAATGTAGTTGAAGAAGTAGTAGAAGAAACCCAG	649
QY	470	TAGATGATAGATTTGACAGACCTGGAACCTGAGTATCTCATGAAGAAACCGACATAGTT	529
Db	650	TAGTTTGAAGAGAAATGATAGAAGAACTCCTGTAGTTGAAGAAGTAGTAGAAGAAACAC	709
QY	530	ACCAGCTGGAAGACACAGTTTTCACAGACTGTAATCAGGATATGGAAGAGAGATGATCTG	589
Db	710	CAGTAGTTGAAGAGAGAGTTGTAGAACAAACACCCAGTAGTTGGAAGAGAAATGGT	765
QY	590	AGCAGGAAAATCCAGATTCCAGTGAAACCCAGTAGTAGAAGATGAAAGATTGCACCATGATA	649
Db	765	-AGAGAAACACCCAGTAGTGGAGAAAGATGTAGTAGAAGAAACACCCAGTAGTTGAAGAAG	823
QY	650	CAGATGATGTAACATACCAAGTCTATCAGAGAACACGACGATATATGAACCTCTGAGAAATG	709
Db	824	AAGTTGTTTGAAGAAACACCCAGTATTTGAAGAAGAAGTTGTAGAAGAAACATCATGATGTTG	883
QY	710	AAGSGATAGAATCACAGAGTAGTAACTCCTCCCTCAGGAGATAATCCTGTAGAGATTCCAC	769
Db	884	AAG--AAGAATGATAGAAGAAACACCCAGTAGTTGAAGAAAAGTAGTAGAAGAAACAC	940
QY	770	AGGTAATTTGTAGAAGAAGT 788	
Db	941	CAGTAGTGGAGAAAAAGT 959	

RESULT 7

ID	Q87589	Q87589 standard; DNA; 585 BP.
AC	Q87589;	
DT	19-DEC-1995	(first entry)
DE	DNA encoding Leucocytozoan protozoa structural protein epitope.	
KW	leucocytozoan protozoa; structural protein; epitope; vaccine; fowl;	
KW	leucocytozoanosis; treatment; ss.	
OS	Leucocytozoan protozoa sp.	
PN	J0708995-A.	
PD	04-APR-1995.	
PF	10-SEP-1993;	226078.
PR	10-SEP-1993;	JP-226078.
PA	(DOBU-) DOBUTSUO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.	
PA	(NISS-) NISSEIKEN KK.	
DR	WPI: 95-167252/22.	
PT	Immune inducing polypeptide against Leucocytozoan protozoa - useful in production of vaccines for treatment of leucocytozoanosis in fowl.	
PS	Claim 1; Page 16; 20pp; Japanese.	
CC	Q87587-89 encode polypeptides having a whole or partial epitope of a structural protein of Leucocytozoan protozoa (see R70491-93). The polypeptides and DNA encoding them are useful in the production of vaccines for the treatment of leucocytozoanosis of fowl.	
CC	Sequence 585 BP; 267 A; 57 C; 155 G; 106 T;	

Query Match	2.8%	Score 68.6;	DB 1;	Length 585;
Best Local Similarity	50.9%	Pred. NO. 6.4e-07		
Matches 254:	Conservative	0: Mismatches 224:	Indels 21:	Gaps 3:

[illegible]

Db 239 TAGTTGAAGAAGAAATGATAGAGAAACCTCCCTGATGTTGAAGAAGTAGTAGAAGAACAC 298
 QY 530 ACCAGCTGGAAGAGACAGTTTTTCCACAGAGCTGTAATCAGGATATGGAAGAGATGATGCTG 589
 Db 299 CAGTAGTTGAAGAAGAGTTGTAGAGAGAACACCACTAGTGGAGAGAAATGGT----- 354
 QY 590 ACCAGGAAATCCAGATTTCCAGTGAACCACTAGTAGTAGAAGATGAAGATGCACCATGATA 649
 Db 354 -AGAAGAAACACCACTAGTGGAGAGAGATGTAGTAGAAGAAACACCACTAGTTGAAGAG 412
 QY 650 CAGATGATGATACATACCAAGTCTATGAGGAACAACCACTATATGAACCTCTAGAAAATG 709
 Db 413 AGTTGTTGAAGAAACACCACTATTTGAAGAGAACTTTAGAGAAACATCATCAGTAGTTG 472
 QY 710 AAGGGATAGAAATCAGAGAACTGCTCCCTCAGGATATCTCTGTAGAGAGATTCCAC 769
 Db 473 AG-----AGAAATGATAGAGAAACACCACTAGTTGAAGAAAGTAGTAGAAGAAACAC 529
 QY 770 AGGTAATGTTAGAGAAAGT 788
 Db 530 CAGTAGTGAAGAAAAAGT 548

RESULT 8

ID Q03875 standard; DNA; 3095 BP.
 AC Q03875;
 DT 24-AUG-1990 (first entry)
 DE Sequence encoding carboxylic terminal part of native GLURP.
 KW Plasmodium falciparum; antigen; malaria; vaccine; GLURP; ss.
 OS Plasmodium falciparum.
 FH key
 FT cds
 FT 1, 2352
 FT /tag= a
 FT /product=GLURP
 FN W09022811-A.
 PD 22-MAR-1990.
 PF 18-SEP-1989; 00218.
 PR 03-MAR-1989; US-218985.
 PR 03-MAR-1989; DK-005191.
 FA (STAT-) Statens Seruminst.
 PI Driegel M, Borre M, Jepsen S, Vuust J, Rieneck K, Wind A, Jakobsen PH;
 DR WPI; 90-115998/15.
 DR P-PSDB; R05804.
 PT Polypeptide(s) derived from Plasmodium falciparum antigen - used in
 PT vaccines and in production of antibodies, for diagnosis and
 PT therapy of malaria.
 PS Disclosure; Fig 7; 108pp; English.
 CC An open reading frame of 2349 bps extends from the 5' terminal end of the
 CC insert to a "TAA" stop codon. It is longest ORF found in the sequence.
 CC Sequence displays some of the characteristics of other malaria nucleic
 CC acid sequences: tandemly repeated motifs, high AT content and a
 CC corresponding preference for codons containing these bases, and a high
 CC content of codons for glutamate. Three major repetitive sequences are:
 CC one motif from bp 34 to bp 156 is repeated from bp 289 to bp 411; another
 CC motif from bp 477 to bp 521 is repeated tandemly twice from bp 522 to bp
 CC 566 and from bp 567 to bp 611; a third motif from bp 1174 to bp 1233 is
 CC repeated tandemly 11 times. This last repetitive region consists of 360bp
 CC repeats differing only in 3 bases GAT coding for aspartate. This region
 CC is flanked to the 5' terminal od a degenerated 60 bp repeat. GC content
 CC of the coding part of the insert is on average 30%, and of the non-coding
 CC 3' terminal 11%
 SQ Sequence 3095 BP; 1443 A; 300 C; 491 G; 861 T;

Query Match 2.7%; Score 65; DB 1; Length 3095;
 Best Local Similarity 46.8%; Pred. No. 7.9e-06;
 Matches 314; Conservative 0; Mismatches 345; Indels 12; Gaps 3;

QY 168 TGCTGATGCTGATGAGATTTTGTGCTGATGATGATGCCAAAGTTTATTAGGACTTTAAGA 227
 Db 831 TGTGATCTGAGTAAAGAAAAGAAATGTATCTGAAGTTGTTGAAGAAAACAAAA 890

QY 228 GAGATCTACTTTCAGAGCCAGCAGTCCCGCCAGAGAGAGGCTGAGCCACACACTGAGCCCGA 287
 Db 891 TTCACAAGATCAGTTGGAAGAAATTCAGTAAGTGAAGATGAATTTGAAGATGTTTCACAC 950
 QY 288 GGAGCAGGTTTCCCTGTGGAGGAGAGAACCCCAAGATATCGAAGATGAAGCAAAAAGAACAAAT 347
 Db 951 TGAACAATTTAGATTTAGATCATAAAACAGTTGATCCAGAAATAGTAGAAGTTGAAGAAAT 1010
 QY 348 TCAGTCCCTTCTCCATGAAA-----TGGTACAGCGAGAACATGTTGAGGGAGAGACTT 401
 Db 1011 TCCTTCAGAACTACATGAAAAATGAAGTGGCTCATCCAGAAATTTGTTGAAATTTAGGAAGT 1070
 QY 402 GCACAAGAGATGGAGCCCAAGAGAACCAACAGAGAGATGATCAGTTTCTTATGGC 461
 Db 1071 TTTTCTGACCAACAAATCAAAATACGAAATTTCAAGAAATTAATGAAGATGATAAAGTGC 1130
 QY 462 GACTGATGATGATAGATTTTGAGACCCCTGGAACTGGAAGTATCTCATGAGAAACCGA 521
 Db 1131 ACATATTCAGCATGAAATAGTAGAAGTAGAAGTAAGAAATACTTCCAGAGATGATAAAATGA 1190
 QY 522 GCATAGTTACCACGTGGAGAGACAGATTTTCAAGACTGTAATCAGGATATGGAAGAGAT 581
 Db 1191 ANAAGTTGAACATGAAATAGTAGAAGTTGAAGAAATTTCTAGCAGAGATGAAAAATGAAA 1250
 QY 582 GATGCTGAGCAGGAAAAATCCAGATCCAGTGAACCACTAGTAGAAGATGAAAGATTGCA 641
 Db 1251 AGG---TCAACATGAAATAGTAGAGTTGAAGAAATTTCTACCAGAGATGATAAAATGA 1307
 QY 642 CCATGATACAGATGATGTAACATACCAAGTCTATGAGGAACAAGCAGTATATGAACCTCT 701
 Db 1308 AAAAGTCAACATGA---AATAGTAGAAGTTGAAGAAATTTCTACCAGAGATGATAAAATGA 1364
 QY 702 AGAAATGAAGGATAGAAATCACAGAACTGTAAGTCTCCCTCAGGATGATATCTGTGAGA 761
 Db 1365 AAAAGTCAACATGAAATAGTAGAGTTGAAGAAATTTCTACCAGAGATGAAAAATGAAA 1424
 QY 762 AGATTCACAGGTAATTTGAGAAGAAAGTAAGCATTTTCTGTGGAAGAACAGCAGGAGT 821
 Db 1425 AGTTGAACATGAAATAGTAGAAGTTGAAGAAATTTCTACCAGAGATGAAAAATGAAAAGG 1484
 QY 822 ACCACCAGATA 832
 Db 1485 TCAACATGAAA 1495

RESULT 9

ID T34620 standard; DNA; 3337 BP.
 AC T34620;
 DT 12-NOV-1996 (first entry)
 DE P. vivax ESP-1 blood stage antigen coding sequence.
 KW ESP-1; blood stage antigen; diagnosis; malaria; infection;
 KW causative agent; antibody; monoclonal; polyclonal; assay; ds.
 OS Plasmodium vivax (clone PvMB3.3.1).
 FH key
 FT exon
 FT 1, .91
 FT /tag= a
 FT /notes="encodes initial (N-terminal) sequence of
 FT hydrophobic amino acids"
 FT 92..230
 FT /tag= b
 FT intron
 FT /note="contains typical malaria intervening
 FT sequence splice sites"
 FT 231..3197
 FT exon
 FT /tag= c

FT US5532133-A.
 PN 02-JUL-1996.
 PD 02-JUN-1993; 072610.
 PF 02-JUN-1993; US-072610.
 PR (UYNK) UNIV NEW YORK STATE.
 PA Barnwell JW;
 PI WPI; 96-321110/32.
 DR

DR P-PSDB; R98747.
PT Antibodies to Plasmodium vivax blood stage antigens - used to
PT diagnose malaria and to determine whether P. vivax is the species
PT responsible for infection
PS Example 4; Column 15-20; 22pp; English.
CC The present sequence encodes a species-specific Plasmodium vivax
CC malarial antigen, PvESP-1. The gene appears to be missing a small
CC portion of its 5' end. This protein is secreted into the plasma of
CC a susceptible mammalian host after infection. Monoclonal/polyclonal
CC antibodies can be utilised in assays used to diagnose malaria, as well
CC as to determine whether P. vivax is the species responsible for the
CC infection.
SQ Sequence 3337 BP; 1304 A; 467 C; 875 G; 691 T;

Query Match 2.5%; Score 61; DB 1; Length 3337;
Best Local Similarity 44.2%; Pred. No. 6.6e-05;
Matches 299; Conservative 0; Mismatches 375; Indels 3; Gaps 1;

QY 146 TAGAARAACTAGGAACTATGATGCTGATGGTGTGATGAGATTTTGTGATGATGCCA 205
DB 1723 TATTAAACAAATCTCTCAAGATCTGATGACGAGAGGCTGTAAACAGTACCATCAAAAGG 1782
QY 206 AAGTTTATTAGGACTTAAAGAGAGATCTACTTCAGAGCCAGCAGTCCCGCCAGAAAGG 265
DB 1783 AAGCACTGTACAGTACCAAGTGCAGTAGGGCCGCGCAGAAAGTCCCAACGGAAGAT 1842

QY 266 CTGAGCCACACTGACCCGAGGAGCAGGTTCCTGTGGAGGCGAGAACCCCGAATATCG 325
DB 1843 TGATGCAACTCCAAGAGGAGCATTTTCAATTAGAAGAACTGCAGAGTCCAGAGGAAG 1902

QY 326 AAGTGAAGCAAAAGACAATTCAGTCCCTTCTCATGAAATGGTACACGCAAGATG 385
DB 1903 GAGAATTAGTATTAGAGGAGAGGAGGAACCAACGGAAGAGAGCCCAAGAGAGAGC 1962

QY 386 TTGAGGAGAGAGACTTCCAACAAGAGATGACCCACAGGAGAACCCACAAGAGATG 445
DB 1963 CACAGAGAGAGAGTCCAGAGAGAAATTAGAGGCACTCCAGAGCAGATTTCCGAAT 2022

QY 446 ATGAGTTCTTATGGGACTGATGTAGATGATAGATTGAGACCCCTGGAACTGAAGTAT 505
DB 2023 TAGAAGAAACCAACAGGAGAGAGTAGAAGAAACCGTAGAGGGCGAAGAACTCAGAG 2082

QY 506 CTATGAGAAACCGCAGATAGTATACACGTGGAGAGACAGTTTCAAGAGCTGTATC 565
DB 2083 GAGAAGAAAGTGGAGAGGTACCTCAGAAAGTAGAAGA---AGTGGAAAGAGTACCTGCAG 2139

QY 566 AGGATATGGAAGAGATGATGCTGAGCAGGAGAAATCCAGATTCAGTGAACCACTAGTAG 625
DB 2140 AAGTAGAAGTGGAGAGGTACCAAGAAAGTAGAAGAGGTACCCGCAAGTAGAAG 2199

QY 626 AAGATGAAGAGATTGCCACCATGATACAGATGATGTAACATACCAAGTCTATGAGGAACAAG 685
DB 2200 AAGTGGAGAGGTACCAAGAGAGTGGAGAGGTACCAAGAGAGTGGAGAGGTACCAAG 2259

QY 686 CAGTATATGAACCTCTAGAAATGAAGGATAGAAATCAAGAGAGTAACTGCTCCCTG 745
DB 2260 AAGAAGTGGAGAGGTACCAAGAGAGTGGAGAGTGGAGAGAGTGAAGAGAGTACAGG 2319

QY 746 AGGATATCTGTAGAGATTCACAGTAAATTTAGAGAGAGTAAAGTATTTCTCTGG 805
DB 2320 TACCAGCGGTAGTAGAAGTACCAAGCGGTAGTAGAAGAGAGGTGCCAGAGAAG 2379

QY 806 AAGAACAGCAGGAAGTA 822
DB 2380 TAGAAGAGAGAGAA 2396

RESULT 10

Q67190

ID Q67190 standard; DNA; 9636 BP.

AC Q67190;

DT 20-FEB-1995 (first entry)

DE P. falciparum transmission blocking target antigen Pfs230 DNA.
KW Protozoan; transmission blocking target antigen; Pfs230; malaria;
KW vaccine; ss.
OS Plasmodium falciparum.
FH Key Location/Qualifiers
FT cds 149.9553
FT /*tag= a
PN W09417187-A.
PD 04-AUG-1994.
PF 18-JAN-1994; U00547.
PR 29-JAN-1993; US-010409.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (USSH) US SEC DEPT HEALTH.
PI Kaslow DC, Williamson KC;
DR WPI; 94-264101/32.
DR P-PSDB; R57474.
PT New Plasmodium falciparum transmission blocking target antigen -
PT useful in antimalarial vaccines, also related DNA, expression
PT vectors and transformed cells
PS Claim 2; Page 24; 63pp; English.
CC The DNA may be used to express Pfs230 in a host cell. It can also
CC be used in vaccines (by incorporation into viral vectors which are
CC then used to infect host cells) and oligonucleotides derived from it
CC can be used to identify homologous proteins in other spp.
SQ Sequence 9636 BP; 4156 A; 974 C; 1422 G; 3084 T;

Query Match 2.4%; Score 59; DB 1; Length 9636;
Best Local Similarity 44.7%; Pred. No. 0.00028;
Matches 273; Conservative 0; Mismatches 335; Indels 3; Gaps 1;

QY 261 AGAGGTGAGCCACACTGAGCCCGAGGAGCAGGTTCCTGTGAGGCGAGAACCCAGAA 320
DB 979 AGATGATGAGGAGAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1038

QY 321 TATCGAGATGAACCAAGAACAAATTCAGTCCCTTCTCCATGAATGGTACAGCAGA 380
DB 1039 AGAAGAGAGGAGGAGAGAGATATGATGATTATGTTATGAAGAAAGTGGGGATGAAC 1098

QY 381 ACATGTTGAGGAGAGAGCTTGCAACAAGAGATGGACCCACAGAGAGAACCAACAAGA 440
DB 1099 AGAAGACATTCACAGAGGAGACATCAGGAAGTAGGTCTGAAATCTTCAGAGAAAG 1158

QY 441 GGATGATGAGTTCTTATGGCGACTGATGTAGATGTAGATTTGAGACCCCTGGAACTGA 500
DB 1159 TTTTATGATGAGATGAAGATTCTGTAGAAGCAGCGGATGGAGATATGATAAGAGTTGA 1218

QY 501 AGTATCTCATGAAGAACCCGAGCATAGTTACCAGCTGGAAGAGACAGTTTCACAGACTG 560
DB 1219 CGAATATTTAGAACCAAGATGGTGATCTTATGATAGTACAAATAAATAATGAAGTGT 1278

QY 561 TAATCAGGATATGAAGAGATGATGCTGAGCAGGAGAAATCCAGATTCAGTGAACCACTG 620
DB 1279 AGATCAAGAGTAGTGAAGAGTAGTGAAGAGTAGTGAAGAGTAGTGAAGAGTAGTGAAGAG 1338

QY 621 AGTAGAAGTGAAGAGATTGCCACCATGATACAGATGATGTAACATACCAAGTCTATGAGGA 680
DB 1339 AGGTGAAGAGTAGTGAAGAGTAGTGAAGAGTAGTGAAGAGTAGTGAAGAGTAGTGAAGAG 1398

QY 681 A---CAAGCAGATATGAACCTCTAGAAATGAAGGATAGAAATCAAGAGTAAGTGC 737
DB 1399 AGGTGAAGAGTAGTGAAGGAGTGAAGAGTAGTGAAGAGTAGTGAAGAGTAGTGAAGAGGT 1458

QY 738 TCCCTCTGAGATATCTGTAGAGATTACAGAGTATTCAGAGAGTATTCAGAGAGTAAAGCATTTT 797
DB 1459 AGGTGAAGAGAGAGGTGAATATGTAGATGAGAGAGAGAGAGAGAGGTGAATATATATCCATT 1518

QY 798 TCCTGTGGAAGACAGCAGGAGAGTACCACCATGATCTTAAAGCTTCAAAAAGAGTGCCTCC 857
DB 1519 TGGTGTGAGAGAGAGAGAGAGAGTGAAGGTGGAGAGTGTTCACCTATGAGAGAGAGGT 1578

QY 858 TACCACACAG 868
| | | | |

DE Human p160 cDNA 160.2.
 KW p160; p62; cytoplasmic; T cell; B cell; development; activation;
 KW modulation; cellular response; cell proliferation; autoimmune disease;
 KW p56-lck; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT 439..3156
 FT CDS /*tag= a
 FT /product= p160
 FT /note= "160.2"
 PN WO9722255-A1.
 PD 26-JUN-1997.
 PF 11-DEC-1996; U19944.
 PR 19-DEC-1995; US-574959.
 PA (DAND) DANA FARBER CANCER INST INC.
 PI Jung I, Shin J, Strominger JL, Vladlamudi RK;
 DR WPI; 97-341351/31.
 DR P-PSDB; W31186.
 PT cDNA encoding p62 and p160 and corresponding proteins - used in the
 PT treatment of autoimmune disease and for T and B cell proliferation,
 PT e.g. for treatment of tumours
 PS Claim 82; Fig 10; 175pp; English.
 CC This cDNA sequence encodes a novel p160 (160.2) which is capable
 CC of activating transcription of a variety of genes upon activation of p62
 CC and is capable of binding to the p62/p56lck complex to modulate Lck
 CC function in a manner similar to p62. The genes transcribed in response to
 CC p160 activation likely include those of which are involved in T or B cell
 CC development/differentiation, T or B cell activation or production of T or
 CC B cell specific factors e.g. lymphokines or antibodies. This p160
 CC polypeptide is also a substrate for serine/threonine kinase activity.
 CC p160 polypeptides can modulate degradation of cellular proteins e.g. cell
 CC cycle regulatory proteins stimulating expression of cell cycle dependent
 CC kinase inhibitors and arresting cell cycle progression at specific
 CC boundaries to thereby modulate cell proliferation. As p160 boosts B cell
 CC response it may be used to treat disorders where this is beneficial, e.g.
 CC infections by pathogenic microorganisms. p160 can be used to expand T
 CC cell populations for treating infectious diseases or cancer and p160
 CC inhibitors could reduce B or T cell responses and may be used to treat a
 CC variety of autoimmune diseases, e.g. diabetes mellitus, arthritis,
 CC multiple sclerosis allergic reactions, Crohn's diseases etc.
 SQ Sequence 3211 BP; 649 A; 990 C; 948 G; 624 T;

Query Match 2.2%; Score 53.4; DB 1; Length 3211;
 Best Local Similarity 47.7%; Pred. No. 0.0035;
 Matches 156; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 271 CCACACACTGAGCCCGAGGAGCAGGTTCTCTGTGGAGGAGACCCCGAGAAATATCGAAGAT 330
 DB 2305 CCACCCCGAGTTGGTCCCTGAGGGGACTCTGTGGGGAGGAGGCCCCCGCCCTGGAAGAG 2364
 QY 331 GAAGCAAAAGACAAATTCAGTCCCTTCTCCATGAATGGTACAGCGCAGAACATGTTGAG 390
 DB 2365 GATTTGACAGTTATTAAATATCAACAGCAGTGATGAAGAGGAGGAGGAAGAGAG 2424
 QY 391 GGAGAAAGACTTGCACACAGAGATGGACCCACGAGGAGACCCACACAGAGGATGATGAG 450
 DB 2425 GAAGAAG 2484
 QY 451 TTTCTTATGGGACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 510
 DB 2485 GAAGACTTTGAGGAAG 2544
 QY 511 GAAGAAACCGAGCATAGTTACCCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 570
 DB 2545 GAAGAGAGTTGAGGAGAGATTTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2604
 QY 571 ATGGAAGAGATGATCTCTGAGCAGGAA 597
 DB 2605 GAGGAGGATGAGGAGGAGAGAGAGAA 2631

RESULT 14

Q21769
 ID Q21769 standard; DNA; 616 BP.
 AC Q21769.
 DT 23-JUL-1992 (first entry)
 DE Sequence encoding Plasmodium falciparum immunogenic peptide IIIC.
 KW Immunogen; vaccine; malaria; epitope; probe; antibody; ss.
 OS Plasmodium falciparum.
 FH Key Location/Qualifiers
 FT 2..616
 FT CDS /*tag= a
 PN WO9203552-A.
 PD 05-MAR-1992.
 PF 14-AUG-1991; F00667.
 PR 14-AUG-1990; FR-010363.
 PA (INSP) INST PASTEUR.
 PI Muller-Hill B, Kun J, Schreiber M, Gysin J, Pereira Da Silva L,
 PI Breton C;
 DR WPI; 92-096896/12.
 DR P-PSDB; R21614.
 PT New polypeptide(s) inducing protective antibodies - having
 PT mediator function on infected red corpuscles and useful as
 PT diagnostic agents and vaccines
 PS Claim 6; Fig 3; 66pp; French.
 CC The polypeptides of the invention contain at least one sequence
 CC having at least one epitope characteristic of a protein present on
 CC the surface of red blood cells (RBC) infected with Plasmodium
 CC falciparum (P.f.). They are useful as immunogens in protective
 CC vaccines against malaria, and for in vitro immuno-detection of P.f.
 CC in tissues or biological fluids. Nucleic acid sequences encoding
 CC the polypeptides are also claimed, and so are nucleotide probes
 CC contg. all or part of the nucleic acid sequences. 144 T;
 SQ Sequence 616 BP; 272 A; 46 C; 154 G;

Query Match 2.0%; Score 49.2; DB 1; Length 616;
 Best Local Similarity 47.0%; Pred. No. 0.017;
 Matches 187; Conservative 0; Mismatches 208; Indels 3; Gaps 1;

QY 339 AGACAATAATTCAGTCCCTTCTCCATGAATGGTACAGCGAGAACATGTTGAGGGAAGA 398
 DB 184 AGGACAACAAGAAATCTGTTACTGAAGAAATAGTAGTGAGAGGATCATTTACTCAAGA 243
 QY 399 CTTGCAACAAGAAGATGGACCCACAGGAGAACACACAAAGAGGATGATGTTCTTAT 458
 DB 244 CATAGTAGAGGAAGAAGATCATGTTACGGAAGAAATAGTAGTGATGAAGATCGGTTAC 303
 QY 459 GCGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 518
 DB 304 TAAGAAATAGTAGAGGATGAAGAAATAGTTACTGAAGAAATAGTAGAGGATGAAGATC 363
 QY 519 CGAGCATAGTTTACCACGTGGAAGAGACAGTTTTCACAAGACTGTAATCAGGATGGAAGA 578
 DB 364 TTTTACTGAGAGAGTTATAGAGAGAAAGATCATTAATTGAGAGAGTTGAAGATGAAGACCA 423
 QY 579 GATGATGCTGTAGCAGGAGAAATCCAGATTCAGTGAACAGTAGTAGAAGATGAAGATTT 638
 DB 424 GTTGCTTGAAGAAAGAAGAGGATCAGTT---ATTAAAGAAATTAATTGACGAAAAATCACT 480
 QY 639 GCACCATGATACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 698
 DB 481 TACTGAAAAAATAGTTGAGAGAGAGAAATCAGTTACTGAGAGAGTTGAAGAGAAAGAAATC 540
 QY 699 TCTAGAAATGAGGATGAGAAATCAGAGAGTAACTG 736
 DB 541 GGTGAGGAGAGAGTAGAGAGAGAGAGAGAGATTTAGTTGTTG 578

RESULT 15

ID N90127
 AC N90127 standard; DNA; 839 BP.
 DT 1-NOV-1989 (first entry)
 DE DNA portion of Plasmodium falciparum rhoptry antigen Ag512

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 1999, 23:10:54 ; Search time 489.51 Seconds
(without alignments)
7609.827 Million cell updates/sec

Title: US-09-040-485-1

Perfect score: 2442

Sequence: 1 CGGGAGCTTGAAGGACACAA.....GGTTAACTTTAAATATTTTT 2442

Scoring table: IDENTITY_NUC

Searched: 2002476 seqs, 76271212 residues

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: gb_est1:*
11: gb_est2:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: gb_est6:*
16: gb_est7:*
17: gb_est8:*
18: gb_est9:*
19: gb_est10:*
20: gb_est11:*
21: gb_est12:*
22: gb_est13:*
23: gb_est22:*
24: gb_est14:*
25: gb_est15:*
26: gb_est16:*
27: gb_est17:*
28: gb_est18:*
29: gb_est19:*
30: gb_est20:*
31: gb_est21:*
32: em_est10:*
33: em_est11:*
34: em_est12:*
35: em_est13:*
36: em_est14:*
37: em_est15:*
38: em_est16:*
39: em_est17:*
40: em_est18:*
41: em_est19:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	489	20.0	851	15	AA203198
2	455.4	18.6	502	12	N31186

C	3	439.4	18.0	600	30	AI084624
	4	432.8	17.7	436	24	AA424475
	5	385.4	15.8	424	15	AA037341
C	6	372.8	15.3	585	15	AA176260
	7	365.8	15.0	574	31	AI267311
C	8	342.8	14.0	598	12	N22713
	9	340.8	14.0	468	10	T69872
	10	340.8	13.9	540	13	W03046
	11	335	13.7	350	10	T07611
	12	333.4	13.7	347	12	N26443
C	13	326	13.3	471	22	AA602346
C	14	325.4	13.3	452	26	AA788915
C	15	325.4	13.3	474	27	AA827544
C	16	316	12.9	575	13	W44451
	17	312.6	12.8	482	26	AA767510
	18	307.4	12.6	392	10	T48492
	19	307.4	12.6	392	20	HUMHBC2999
C	20	296.8	12.2	598	12	N48349
	21	294.6	12.1	404	11	H26237
	22	294.4	12.1	315	19	C00261
	23	293.6	12.0	549	15	AA176259
	24	290.8	11.9	379	12	H81879
C	25	281	11.5	290	22	AA593699
C	26	279	11.4	426	24	AA157848
C	27	276.6	11.3	361	25	AA669808
C	28	270	11.1	417	22	AA576235
	29	267	10.9	289	17	AA304742
C	30	263	10.8	405	25	AA702708
	31	253	10.4	279	15	H96029
	32	250.4	10.3	265	26	AA745282
C	33	250.2	10.2	397	25	AA079718
C	34	250	10.2	404	23	AI290103
C	35	246.8	10.1	410	21	AA541348
C	36	245.8	10.1	290	11	H25536
C	37	243.8	10.0	446	15	AA159372
C	38	239.4	9.8	412	12	H99385
C	39	234.6	9.6	454	21	AA523182
C	40	227.4	9.3	391	12	N23024
C	41	226.4	9.3	324	24	AA601990
C	42	215.2	8.8	387	15	H96030
C	43	188.4	7.7	330	12	H99337
C	44	169.4	6.9	285	12	H90076
	45	157	6.4	160	17	AA309066

ALIGNMENTS

RESULT 1

AA203198

LOCUS

DEFINITION z57a02.r1 Soares fetal liver spleen 1NF1S s1 Homo sapiens CDNA
clone 446570 5', mRNA sequence.

ACCESSION AA203198

NID 91798908

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Euthera; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 851)

AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfs, T., Soares, M., Tan, F.,
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.

TITLE The WashU-Merck EST Project

JOURNAL

Unpublished (1995)

COMMENT

Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

24-JAN-1997

Db 538 GAATATTAGTAGTAAAGAACATA 562

||||| || | ||||| |

RESULT 2

N31186 502 bp mRNA EST 10-JAN-1996
LOCUS yx64f08.r1 Homo sapiens cDNA clone 266535 5'
DEFINITION N31186
ACCESSION g1151585
NID g1151585
EST.
SOURCE human clone=266535 primer=T7 library=Soares melanocyte 2NbHM vector=pR73D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) RsiteI=Not I Rsite3=Eco RI Male. 1st strand cDNA was primed with a Not I - oligo(GT) primer [5'-GTGTTACCAATCTGAAGTGGGCGCCGACGTTTTTTTTTTTTTTT-3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pR73 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P. Albino.

ORGANISM Homo sapiens
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 502)

REFERENCE HALLIER,L., CLARK,N., DUBUQUE,T., ELLISTON,K., HAWKINS,M., HOLMAN,M., HULIMAN,M., KUCABA,T., LE,M., LENNON,G., MARRA,M., PARSONS,J., RIFKIN,L., ROHLFING,T., SOARES,M., TAN,F., TREVASKIS,E., WATERSTON,R., WILLIAMSON,A., WOHLDMANN,P. and WILSON,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 331
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
FEATURES Location/Qualifiers
source 1..502
/organism="Homo sapiens"
/clone="266535"
BASE COUNT 165 a 69 c 77 g 190 t 1 others
ORIGIN <1..>502

Query Match 18.6%; Score 455.4; DB 12; Length 502;
Best Local Similarity 98.0%; Pred. No. 5.9e-90;
Matches 492; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

QY 1205 GG GTTTTCATATATTAGTAATTAACACTAACTGTGGACTGACTGTGTACACTGTG 1264
|||||
Db 1 GGT TTTTCATATATTAGTAATTAACACTAACTGTGGACTGACTGTGTACACTGTG 50
|||||

QY 1265 TTAAACATGATTTAAAAGCTATTAAAGAGTACTTTTGTTTAGCAGCTCTTAAAAACGCCTAAC 1324
|||||

Db 61 TTAACATGATTTAAAAGCTATTAAAGAGTACTTTTGTTTAGCAGCTCTTAAAAACGCCTAAC 120
|||||

QY 1325 AGAGATCATCATTAGCTGTGAAGATTTGAGTTGATATATACCTGCAGCTGATATTCTTATCA 1384
|||||

Db 121 AGAGATCATCATTAGCTGTGAAGATTTGAGTTGATATATACCTGCAGCTGATATTCTTATCA 180
|||||

QY 1385 AAAATTCTACATTAGCTTTAAGTTGCAGATTAAACACTTTTTGAAACCCTTTGTAGCTTTT 1444
|||||

|||||
Db 181 AAAATTTCTACATTAGCTTTAGTGTTCAGATTAAACACTTTTGAATTTTGTAGCTTTT 240
QY 1445 AGCTGATTAAATAGAAAAA-TTAATATTTTCAGTGAAGTTTAAATATATCATTTATTTAT 1503
Db 241 AGCTGATTAATTAAGAAAAATTTAATATTTTCAGTGAAGTTTAAATATATCATTTATTTAT 300
QY 1504 TTTTAAATGAGAGGGGAAAGCTGAAATTCCTTGTAAAGACACAAAGGAAAAAGATGGC 1563
Db 301 TTTTAAATGAGAGGGGAAAGCTGAAATTCCTTGTAAAGACACAAAGGAAAAAGATGGC 360
QY 1564 CTTACTATTATCATGCAAAAATGCTTTTGTGGCACCCTCAGATTAAATCAATATAAGCTAT 1623
Db 361 CCTACTATTATCATGCCAAAATGCTTTTGTGGCACCCTCAGATTAAATCAATATAAGCTAT 420
QY 1624 ACTCTCTT-CACATTTGTTAAATTTTGAAGAAACCTGTATAAATTACTGGTGCATAAC- 1682
Db 421 AGCTCTCTCCAGCATTTGTTTAAATTTNAGAAAACCTGTATAAATTACTGGTGCATAACT 480
QY 1682 TTAAGATTATTCTGCCTTTGG 1703
Db 481 TTAAGATTATTCTGCCCTTTG 502

RESULT 3
LOCUS AI084624/c 600 bp mRNA EST 01-OCT-1998
DEFINITION O278c04.x1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA
clone IMAGE:1681446 3', mRNA sequence.
ACCESSION AI084624
NID G3423047
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1338 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 444.
Location/Qualifiers
1..600
/organism="Homo sapiens"
/note="Vector: pT7T3D (Pharmacia) with a modified
polylinker V-TYPE: phagemid; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGCATTTTATTTTATTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo."
/db_xref="taxon:9606"
/clone="IMAGE:1681446"
/clone_lib="Soares_senescent_fibroblasts_NbHSF"
/tissue_type="senescent fibroblast"
/lab_host="DH10B (ampicillin resistant)"

207 a 106 c 74 g 213 t

Query Match 18.0%; Score 439.4; DB 30; Length 600;
Best Local Similarity 99.1%; Pred. No. 1.9e-86;
Matches 453; Conservative 0; Mismatches 1; Indels 3; Gaps 1;
QY 1986 CTATAAGCAAGCCATAAGAAGTGGTTTGATCGATATATTAGGGGTAGCTCTTCAATTTTG 2045
Db 600 CTATAAGCAAGCCATAAGAAGTGGTTTGATCGATATATTAGGGGTAGCTCTTCAATTTTG 541
QY 2046 TTAACATTAAAGTAAGGTGACTTTTCCCTCGCTTTTAGGATTAAATCAAGATACCTT 2105
Db 540 TTAACATTAAAGTAAGGTGACTTTTCCCTCGCTTTTAGGATTAAATCAAGATACCTT 481
QY 2106 CTATATTTTATCATACTATAGATCATAGTATTATATACAAATAGTAGTGAGTCCCTCGATGGTA 2165
Db 480 CTATATTTTATCATACTATAGATCATAGTATTATATACAAATAGTAGTGAGTCCCTCGATGGTA 421
QY 2166 CTCGATGTGTAATGAAACCTGAAATAATAATAGATAATAAGAAAAGCAATAATTTTCTA 2225
Db 420 CTCGATGTGTAATGAAACCTGGA---AATAATAAGATAATAAGAAAAGCAATAATTTTCTA 364
QY 2226 AAGCTGTGCTGTCGGTGATACAGAGATGATCTCAAAATTATAATAAACTCTTCATTTTG 2285
Db 363 AAGCTGTGCTGTCGGTGATACAGAGATGATCTCAAAATTATAATAAACTCTTCATTTTG 304
QY 2286 TGAATTATAGAAGCTACTTTTATAAGCCATATTTTATAGGGAACCTAAGGAGTGACA 2345
Db 303 TGAATTATAGAAGCTACTTTTATAAGCCATATTTTATAGGGAACCTAAGGAGTGACA 244
QY 2346 TAGAACTGATGAATGAGTAAAGTAAGTTTGTCTGGATTTTGTAGAACTCTGGACGTTG 2405
Db 243 TAGAACTGATGAATGAGCAAAAGTAAGTTTGTCTGGATTTTGTAGAACTCTGGACGTTG 184
QY 2406 AGGATTCATTATGCTGTGGTTAACTTTAAATATATTTT 2442
Db 183 AGGATTCATTATGCTGTGGTTAACTTTAAATATATTTT 147

RESULT 4
LOCUS AA424475
DEFINITION zvr82e07.r1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone
760164 5', mRNA sequence.
ACCESSION AA424475
NID G2103427
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-Merck EST Project 1997
Unpublished (1997)

COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 425.
Location/Qualifiers
1..436
/organism="Homo sapiens"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week

FEATURES
source

(total) fetus material with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCGCTAAATTTTCTTTTCTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo.
/db_xref="taxon:9606"
/clone="760164"
/clone_lib="Soares total fetus Nb2HP8 9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
BASE COUNT 147 a 57 c 70 g 162 t
ORIGIN

Query Match 17.7%; Score 432.8; DB 24; Length 436;
Best Local Similarity 99.5%; Pred. No. 5.1e-85;
Matches 434; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1240 GTTGGACTGACTGTGTACACTGTGTTAAACATGATTTAAAGCTATTAAGACTATTG 1299
Db 1 GTTGGACTGACTGTGTACACTGTGTTAAACATGATTTAAAGCTATTAAGACTATTG 60
QY 1300 TGTAGCACCTCTTAAAAACGCTAACAGAGATCATCATTTAGCTGTGAAGATTGAGTTGTA 1359
Db 61 TGTAGCACCTCTTAAAAACGCTAACAGAGATCATCATTTAGCTGTGAAGATTGAGTTGTA 120
QY 1360 TATACCTGCTGATATTTTATCAAAAATTTTACATTTAGCTTTAAAGTGTTCAGATTAA 1419
Db 121 TATACCTGCTGATATTTTATCAAAAATTTTACATTTAGCTTTAAAGTGTTCAGATTAA 180
QY 1420 CACTTTTGAACCTTTGTAGCTTTTACCTGATTAATTAAGAAAATTAATTTCAAGTAA 1479
Db 181 CACTTTTGAACCTTTGTAGCTTTTACCTGATTAATTAAGAAAATTAATTTCAAGTAA 240
QY 1480 AGTTTAAATTTATCATTTATTTTAAATGAGGGAAGCTGAAATTCCTTGT 1539
Db 241 AGTTTAAATTTATCATTTATTTTAAATGAGGGAAGCTGAAATTCCTTGT 300
QY 1540 TAAGACACAGGAAAAAGATGGCCCTACTATTATCATGCAAAAATGCTTTGTTGGCACC 1599
Db 301 TAAGACACAGGAAAAAGATGGCCCTACTATTATCATGCAAAAATGCTTTGTTGGCACC 360
QY 1600 TCAGATTAATCATATATAGCTATAGTCTCTTCAGCAATTTGTTTAAATTTAGAAAACCT 1659
Db 361 TCAGATTAATCATATATAGCTATAGTCTCTTCAGCAATTTGTTTAAATTTAGAAAACCT 420
QY 1660 GTATAAATTACTGGTG 1675
Db 421 GTATAAATTACTGGTG 436

RESULT 5
LOCUS AA037341 424 bp mRNA EST 25-NOV-1996
DEFINITION zc52d12.r1 Soares senescent fibroblasts NBHSF Homo sapiens cDNA
clone 325943 5', mRNA sequence.
ACCESSION AA037341
NID g1512441
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 424)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hulman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)

COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 815 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 194.
Location/Qualifiers
1. 424
/organism="Homo sapiens"
/polylinker="vector: pT73D (Pharmacia) with a modified
RI: TGTACCAATCTGAAGTGGAGCGCGCTAAATTTTCTTTTCTTTT
3', double-stranded cDNA was size selected, ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT73
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo.
/db_xref="taxon:9606"
/clone_lib="Soares senescent fibroblasts NBHSF"
/tissue_type="senescent fibroblast"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT 121 a 80 c 82 g 139 t 2 others
ORIGIN

FEATURES

source
Query Match 15.8%; Score 385.4; DB 15; Length 424;
Best Local Similarity 98.6%; Pred. No. 1.2e-74;
Matches 419; Conservative 0; Mismatches 3; Indels 3; Gaps 3;
QY 1712 GTAATTCCTCCAGCACTAGAGACCGCTCAGTCTCTTACTAGTAACTCAGTAACGC 1771
Db 1 GTAATTCCTCCAGCACTAGAGACCGCTCAGTCTCTTACTAGTAACTCAGTAACGC 60
QY 1772 CTTGAGCTGGTTCATTGAGGATGTGTGAAAAGCTCACAGAGCCCGATGCTGCTGCTA 1831
Db 61 CTTGAGCTGGTTCATTGAGGATGTGTGAAAAGCTCACAGAGCCCGATGCTGCTGCTA 119
QY 1832 TTTCACGGCAATGAGCCCTTTTCTTCTACACTGAAGATTTTCTTCTTATTAATGTGGT 1891
Db 120 TTTCACGGCAATGAGCCCTTTTCTTCTACACTGAAGATTTTCTTCTTATTAATGTGGT 179
QY 1892 TTATTTGGGCTCAGAAATAATTCCTGTTGAAAATAATCCTTTGTGAGAAAAGAGGT 1951
Db 180 TTATTTGGGCTCAGAAATAATTCCTGTTGAAAATAATCCTTTGTGAGAAAAGAGGT 239
QY 1952 AGCTACACATCATTTTGAAGAGCACTAGCAACTATAAGCAAGCCATAGAAGTGGT 2011
Db 240 AGCTACACATCATTTTGAAGAGCACTAGCAACTATAAGCAAGCCATAGAAGTGGT 299
QY 2012 TTGATCATATATTAGGGTAGCTCTTGAATTTTGTACATTAAGATGAAGTGAATTTT 2071
Db 300 TTGATCATATATTAGGGTAGCTCTTGAATTTTGTACATTAAGATGAAGTGAATTTT 359
QY 2072 CCCCTCTTTTAGGATT-AAAATCAAGATACCT-CTATATTTTATCATATAGATCA 2129
Db 360 CCCCTCTTTTAGGATT-AAAATCAAGATACCT-CTATATTTTATCATATAGATCA 419
QY 2130 TAGTT 2134
Db 420 TAGTT 424
RESULT 6

```

AA176260/c  AA176260  585 bp  mRNA  EST  30-DEC-1996
LOCUS      zp29h01.s1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA
DEFINITION clone 610897 3', mRNA sequence.
ACCESSION  AA176260
NID        91757409
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 585)
AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Ruhlifing,T., Tan,F., Trevaskis,E.,
            Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
TITLE     WashU-Merck EST Project
JOURNAL   Unpublished (1995)
COMMENT   Contact: Wilson RK
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40M3 fwd. from Amersham
            High quality sequence stop: 361.
            Location/Qualifiers
                1..585
                /organism="Homo sapiens"
                /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                XhoI; Cloned unidirectionally. Primer: Oligo dt. NT2
                cells (Ntera-2/c1.d1) induced with Retinoic Acid for 24
                hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5'
                adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor
                sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3"
                /db_xref="taxon:9606"
                /clone_lib="610897"
                /clone_lib="Stratagene neuroepithelium (#937231)"
                /dev_stage="Ntera-2/RA neuroepithelial cells"
                /lab_host="SOLR (kanamycin resistant)"
                complement(<1..>585)
BASE COUNT  193 a 112 c 83 g 189 t 8 others
ORIGIN
mRNA
Query Match 15.3%; Score 372.8; DB 15; Length 585;
Best Local Similarity 91.6%; Pred. No. 6.7e-72;
Matches 413; Conservative 0; Mismatches 34; Indels 4; Gaps 2;
QY 1993 CAAAGCCATAAGAGTGGTTGTATCATATATATAGGGGTAGCTTGTATTTGTTACAT 2052
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 584 CCAANCCATAGGAGNGTGTACCCCAATATTCGGGGTAGCCCTCGATTTGGTTACCAT 525
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2053 TAAGTAAGGTGACTTTTCCCTCTCTTTAGGATTAATAATCAAGATACCTCTATATT 2112
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 524 TAGGATAGGGGGCTTTTCCCTCTCTTTTNGGATNAANAANCAAGGTACTTCTATATT 465
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2113 TTATCACTATAGATCATAGTTATTATACAACTAGTAGTCTCTG-CATGGGTACTCGAT 2171
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 464 TTANCACTATAGATCATAGTTATTATNCAATAGTAGTCTCTGCGCATGGGTACTCGAT 405
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2172 GGTATGAACTGAATAATAATAGATAATAAGAAAGCAATAATTTCTAAAGCTG 2231
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 404 GTGTAATGAACCTGA---AATAATAGATAATAAGAAAGCAATAATTTCTAAAGCTG 348
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2232 TGCTGTGGTGATACAGATGATCTCAATTTATATAAATCTCTTCATTTTGTGAATT 2291
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 347 TGCTGTGGTGATACAGATGATCTCAATTTATATAAATCTCTTCATTTTGTGAATT 288
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

AA176260/c  AA176260  574 bp  mRNA  EST  17-NOV-1998
LOCUS      aq63e03.x1 Stanley Frontal SN pool 2 Homo sapiens cDNA clone
DEFINITION IMAGE:2035612 similar to SW:ASPH_HUMAN Q12797 ASPATYL/ASPARAGINYL
            BETA-HYDROXYLASE ;, mRNA sequence.
ACCESSION  AA1767311
NID        93886478
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 574)
AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE     WashU-NCI human EST Project
JOURNAL   Unpublished (1997)
COMMENT   Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 439.
            Location/Qualifiers
                1..574
                /organism="Homo sapiens"
                /note="Organ: Brain; Vector: pCR2.1-TOPO (Invitrogen);
                Site_1: EcoRI; Total RNA (purified with Trizol and DNaseI
                before use) was reverse transcribed using a modified
                oligo-dT primer containing RsaI and HindIII sites.
                Double-stranded cDNA was digested with RsaI, resulting in
                blunt ended cDNA of an average 0.1-2 kb in length.
                Digested cDNA was split into two sets, one used as is as
                the driver, the other set was split in half again and each
                half linked to a different adaptor
                (5'-TCGAGCGCCGCCGGCAGGT-3' or 5'-
                AGGCGTGGTGGCGAGGCGGT-3'), to be used as tester.
                Subtraction was performed using the Clontech PCR select
                cDNA subtraction kit. Pool of two schizophrénics, male age
                44 and female age 56 (S-116, S-118) subtracted by pool of
                two mentally normal male individuals ages 41 and 53
                (S-124, S-141). Tissues were obtained from the Stanley
                Neuropathology Consortium (www.stanleylab.org). Library
                constructed and subtracted by Dr. Nancy Johnston [(410)
                614-3918, nlj@welchlink.welch.jhu.edu].
                /db_xref="taxon:9606"
                /clone="IMAGE:2035612"
                /clone_lib="Stanley Frontal SN pool 2"
                /tissue_type="frontal lobe (see description)"
                /lab_host="DH10B (phage-resistant)"
BASE COUNT  164 a 150 c 109 g 150 t 1 others
ORIGIN

```


TITLE WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 338
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES

source
1..468
/organism="Homo sapiens"
/clone="81068"

BASE COUNT 159 a 90 c 114 g 102 t 3 others
ORIGIN

Query Match 14.0%; Score 340.8; DB 10; Length 468;
Best Local Similarity 91.6%; Pred. No. 6.5e-65;
Matches 416; Conservative 0; Mismatches 29; Indels 9; Gaps 5;

QY 472 GATGATAGATTGGACCCCTGGACCTGAAGTATCTCATGAAGAAACCGACGATAGTTAC 531
|||
Db 6 GAGGATAGATTGGACCCCTGGACCTGAAGTATCTCATGAAGAAACCGACGATAGTTAC 65
QY 532 CACGTGAAGACAGAGTTTCACAAAGACTGTAATCAGGATATGGAAGAGATGCTCTGAG 591
|||
Db 66 CACGTGAAGACAGAGTTTCACAAAGACTGTAATCAGGATATGGAAGAGATGCTCTGAG 125
QY 592 CAGGAAATCCAGATTCACAGTGAACCTAGTAGAAGATGAAAGATTCACCATGATACA 651
|||
Db 126 CAGGAAATCCAGATTCACAGTGAACCTAGTAGAAGATGAAAGATTCACCATGATACA 185
QY 652 GATGATATACATACCAAGTCTATGAGAACCAAGCAGTATATGAACCTCTAGAAATGAA 711
|||
Db 186 GATGATATACATACCAAGTCTATGAGAACCAAGCAGTATATGAACCTCTAGAAATGAA 245
QY 712 GGGATACAAATCACAGAGTAACCTGCTCCCTCC-TGAGGATATCTGTAGAGATTCACA 770
|||
Db 246 GGGATACAAATCACAGAGTAACCTGCTCCCTCCCTGTGAGGATATCTGTAGAGATTCACA 305
QY 771 GGTAAATGTGTA-GAAGAAAGTAGCAATTTTCTCTGT-GGAAGAACACAGCAGGAAGTACCACCA 828
|||
Db 306 GGTAAATGTAGGACAGTAGCAATTTTCTCTGTGAGGAGAACACAGCAGGAAGTACCACCA 365
QY 829 GATACCTT--AAAGCTTCAAAAGACTGCC--CTACCACACAGGAGGACCGCCTTAA 882
|||
Db 366 GNTACTTTAAAGCTTCAAAAGAGATGCCCCCTACCACACAGGNGGGACCGCCTTAA 425
QY 883 CCATACCTCCAAAGATGGCTGTGATAGATCTT 916
|||
Db 426 CCTACGGTTTCAAAAGGTGGCTGTGTGTGCTTT 459

RESULT 10

W03046 540 bp mRNA EST 18-APR-1996
LOCUS za04d08.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone 291567
DEFINITION 5', mRNA sequence.
ACCESSION W03046
NID g1275093
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 540)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 215.

FEATURES

source

1..540
/organism="Homo sapiens"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCGCAGCTTTTCTTTTCTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M. Fatima Bonaldo. RNA from normal foreskin melanocytes
(FS374) was kindly provided by Dr. Anthony P. Albino."
/db_xref="taxon:9606"
/clone="291567"
/clone_lib="Soares melanocyte 2NbHM"
/sex="Male"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT 157 a 93 c 124 g 163 t 3 others
ORIGIN

Query Match 13.9%; Score 340; DB 13; Length 540;
Best Local Similarity 91.3%; Pred. No. 9.9e-65;
Matches 439; Conservative 0; Mismatches 28; Indels 14; Gaps 7;

QY 1509 TAAATGAGAGGGGAAGCTGAAATTCCTTGTAAACACACAAAGAAAGAAATGGCCCTAC 1568
|||
Db 1 TAAATGAGAGGGGAAGCTGAAATTCCTTGTAAACACACAAAGAAAGAAATGGCCCTAC 60
QY 1569 TATTATCATGCAAAATGCTTTGTTGGCACCCTCAGATTAAATCATATATAGCTATAGTCT 1628
|||
Db 61 TATTATCATGCAAAATGCTTTGTTGGCACCCTCAGATTAAATCATATATAGCTATAGTCT 120
QY 1629 CTTGAGCATTTGTTTAAATTTAGAAAACCTGTATAAATTACTGGTGCATACCTTAAAGA 1688
|||
Db 121 CTTGAGCATTTGTTTAAATTTAGAAAACCTGTATAAATTACTGGTGCATACCTTAAAGA 180
QY 1689 TTATTCTGCTTTGGCTAATTGAGTAATTCCTCCCTCCAGCAGTAGACCGCTCAGTGCCTC 1748
|||
Db 181 TTATTCTGCTTTGG-TAATTGAGTAATTCCTCCCTCCAGCAGTAGACCGCTCAGTGCCTC 239
QY 1749 TTACTAGATGAACCTCAGTAACGCTTTGAGCTGGGTTGATTGAGGATGTGTGAAAAAGCTC 1808
|||
Db 240 TTACTAGATGAACCTCAGTAACGCTTTGAGCTGGGTTGATTGAGGATGTGTG-AAGAGCTC 298
QY 1809 ACAGAGCCCGATGCTGCTGTATTTTCAC-GGCATGAGCCCTTTTCTTCTACACTG-- 1866
|||
Db 299 ACAGAGCCCGATGCTGCTGTATTTTCACGGGCAATGAGCCCTTTTCTTCTACACTGGA 358
QY 1866 AGATTTTCTTCTTATTAATGCTGCTTTATTTT---GGCTCAGAAAATAATGCTCTGT 1921
|||
Db 359 AGATTTTCTTCTTATTAATGCTGCTTTATTTTGGGGCTCAGGAAAATAATGCTCTGT 418

Db	241	GAGTGTGATATACCTGCACTGACTGATATCTTATCAAAAAATTTCTACATTAGCTTTTAAGTGTT	300
Qy	1412	CAGATTACACACTTTTGAACACTTTTGTAGCTTTTGTAGCTGATTAAT	1456
Db	301	CAGATTACACACTTTTGAACACTTTTGTAGCTTTNAGCTGGATTAAT	345
RESULT	13		
AA602346/c			
LOCUS	AA602346	471 bp	24-SEP-1997
DEFINITION	nc089n06.s1 NCI_CGAP_AAL Homo sapiens cDNA clone IMAGE:1114043,	EST	mrna
ACCESSION	sequence.		
NID	AA602346		
KEYWORDS	G2436324		
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;		
TITLE	Primates; Catarrhini; Hominiidae; Homo.		
JOURNAL	1 (bases 1 to 471)		
COMMENT	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
	Tumor Gene Index		
	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert_Strausberg@nih.gov		
	Tissue Procurement: L. Jeffrey Medelros, M.D., Michael R.		
	Emmert-Buck, M.D., Ph.D.		
	cDNA Library Preparation: Stratagene, Inc., David B. Krizman,		
	Ph.D.		
	cDNA Library Arraying: Greg Lennon, Ph.D.		
	DNA Sequencing by: Washington University Genome Sequencing Center		
	Clone distribution: NCI-CGAP clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	www-bio-llnl.gov/bbrp/image/image.html		

```

Seq primer: -40ml3 fwd. ET from Amersham.
Location/Qualifiers
1. .471
/organism="Homo sapiens"
/notes="Organ: adrenal gland; Vector: Bluescript SK-;
Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally.
Primer: Oligo df. Two pooled bulk adrenal adenomas. 5'
adaptor sequence: 5' GAATTCGGCAGAG 3' 3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTT 3' Average insert
size: 1.6 kb."
/db_xref="taxon:9606"
/clone="IMAGE:1114043"
/clone_lib="NCI_CGAP_AAL"
/tissue_type="adrenal adenoma"
/lab_host="SOLR (kanamycin resistant)"
/lab 83 c 56 g 164 t

BASE COUNT 168 a
ORIGIN

Query Match 13.3%; Score 326; DB 22; Length 471;
Best Local Similarity 99.1%; Pred. No. 1.1e-61;
Matches 339; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 2101 TACTCTCATATTTTATCAGCTATAGATCATAGTTATTATACAATGTAGTGAGTCTCTGCAT 2160
|||||
Db 471 TACTCTCATATTTTATCAGCTATAGATCATAGTTATTATACAATGTAGTGAGTCTCTGCAT 412
|||||

QY 2161 GGGTACTCGATGTGTAATGAAACCTGAAATTAATAAGATAATAAGAAAGCAATAATT 2220
|||||
Db 411 GGGTACTCGATGTGTAATGAAACCTGA---AATAATAAGATAATAAGAAAGCAATAATT 355
|||||

QY 2221 TTCTAAAGCTGTGCTGCGGTGATACAGAGATGATACTCAATTTATATAAACTCTTCA 2280
|||||
Db 354 TTCTAAAGCTGTGCTGCGGTGATACAGAGATGATACTCAATTTATATAAACTCTTCA 295
|||||

```

Qy	2281	TTTTGTGAATATATAGAAGCTACTTTTTTATATAAGCCATATATTTTTTAGGAACTAAGGAG	2340
Db	294	TTTTGTGAATATATAGAAGCTACTTTTTTATATAAGCCATATATTTTTTAGGAACTAAGGAG	235
Qy	2341	TGACATAGAAGCTGATGATGAGTAAAGTAAAGTATTTTCTGGATTTTGTAGAACTCTGGA	2400
Db	234	TGACATAGAAGCTGATGATGAGTAAAGTAAAGTATTTTCTGGATTTTGTAGAACTCTGGA	175
Qy	2401	CGTTGAGGATTCATTATGCTGCTGGTTAACTTTTAATAATATTTT	2442
Db	174	CGTTGAGGATTCATTATGCTGCTGGTTAACTTTTAATAATATTTT	133
RESULT	14		
AA788915/c			
LOCUS	AA788915	452 bp	mRNA EST 06-FEB-1998
DEFINITION	ae93304.s1	Stratagene schizo brain s11	Homo sapiens cDNA clone
ACCESSION	AA788915		1020355.3, mRNA sequence.
NID	92849035		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 452) Hillier L., Allen M., Bowles L., Dubuque T., Geisel G., Jost S., Krizman D., Kucaba T., Lacy M., Le N., Lennon G., Marra M., Martin J., Moore B., Schellenberg K., Steptoe M., Tan F., Theisinger B., White Y., Wyllie T., Waterston R. and Wilson R.		
TITLE	WashU-NCI human EST Project		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -40m13 fwd. Et from Amersham High quality sequence stop: 451. Location/Qualifiers 1. .452 /organism="Homo sapiens" /note="Vector: Bluescript SK-; Site_1: EcoRI; Library constructed from S-11 frontal lobe, male, 34 years old, 50% caucasian, 50% Aleutian. Schizophrenic suicide. Random primed into EcoRI site of ZAP II Vector. Mass excised. Custom library. Avg insert length 1.4kb. Material obtained by Johnston N., Torrey, E.F., Yolken R., and the Stanley Neuropathology Consortium - Analysis of RNAs from the Brains of Individuals with Psychiatric Diseases (Unpublished) Stanley Neurovirology Laboratory, Johns Hopkins School of Medicine, Baltimore MD." /db_xref="taxon:9606" /clone="1020353" /clone_lib="Stratagene schizo brain s11" /sex="male" /tissue_type="schizophrenic brain S-11 frontal lobe" /dev_stage="34 years old" /lab_host="SOLR (kanamycin resistant)"		
BASE COUNT	163 a	81 c	55 g 153 t
ORIGIN			
Query Match	13.3%;	Score 325.4;	DB 26; Length 452;
Best Local Similarity	98.8%;	Pred. No. 1.5e-61;	
Matches 33;	Conservative	0; Mismatches	1; Indels 3; Gaps 1;
QY	2100	ATATCTCTATATTTTATCATATAGATCATAGTATTATATACATATGATGAGCTCTGCA	2159

M P S R C H _ p p

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu May 13 15:27:35 1999; MasPar time 2.53 Seconds
Tabular output not generated.
63.976 Million cell updates/sec

Title: >US-09-040-485-6
Description: (1-10) from US09040485.pep
Perfect Score: 64
Sequence: 1 APPEDNPVED 10

Scoring table: PAM 150
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29

Statistics: Mean 14.497; Variance 50.537; scale 0.287

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				Pred. No.	
Result No.	Score	Query Match	Description	ID	
1	48	75.0	Mouse inhibitor of ap	W19749	9.40e+01
2	45	70.3	Retinoblastoma (RB) p	R36534	1.84e+02
3	45	70.3	Retinoblastoma p110R	R71680	1.84e+02
4	45	70.3	Recombinant p110RB p	R71681	1.84e+02
5	45	70.3	Retinoblastoma suscep	W09411	1.84e+02
6	45	70.3	Retinoblastoma tumour	R74271	1.84e+02
7	45	70.3	Cancer suppressing gen	R05305	1.84e+02
8	45	70.3	Predicted retinoblast	R06289	1.84e+02
9	45	70.3	RAS associated GAP NF	R59922	1.84e+02
10	45	70.3	RAS associated GAP NF	R59921	1.84e+02
11	45	70.3	Human neurofibromin.	W13280	1.84e+02
12	45	70.3	Nfl gene product.	R22288	1.84e+02
13	44	68.8	Vibrio cholerae strai	R86553	2.30e+02
14	44	68.8	Vibrio cholerae El To	R86555	2.30e+02
15	44	68.8	Truncated form of hum	R08390	2.30e+02
16	44	68.8	Vibrio cholerae polyP	R72856	2.30e+02
17	44	68.8	Incomplete form of hu	R07661	2.30e+02
18	44	68.8	Sequence encoded by v	R08391	2.30e+02

19 44 68.8 979 2 R08338 Complete form of huma 2.30e+02
20 43 67.2 510 4 R23365 SIVmac239 gag gene pr 2.87e+02
21 42 65.6 506 1 P80802 Sequence encoded by g 3.57e+02
22 42 65.6 521 1 P80807 Sequence of gag prote 3.57e+02
23 41 64.1 14 18 R38376 Potato debranching en 4.44e+02
24 41 64.1 14 13 R64987 Amylopectin starch de 4.44e+02
25 41 64.1 164 4 R23310 TNF polypeptide mutel 4.44e+02
26 41 64.1 360 13 R71677 Human CSBP2. 4.44e+02
27 41 64.1 360 13 R71676 Human CSBP1. 4.44e+02
28 41 64.1 403 23 W18084 Human Aurora-2. 4.44e+02
29 41 64.1 499 8 R42122 NK-1 cellulase. 4.44e+02
30 41 64.1 520 17 R92156 3'-Hydroxymethylcephe 4.44e+02
31 41 64.1 592 2 R05227 Sequence of rat C kin 4.44e+02
32 41 64.1 823 1 P80972 Sequence encoded by 3 4.44e+02
33 41 64.1 823 1 P81502 delta-endotoxin again 4.44e+02
34 41 64.1 1163 14 R85147 CryIC/CryIA(B) chimera 4.44e+02
35 41 64.1 1176 29 W46737 Amino acid sequence o 4.44e+02
36 41 64.1 1176 28 W47035 Bacillus thuringiensis 4.44e+02
37 41 64.1 1189 13 R71463 Crystall protein encod 4.44e+02
38 41 64.1 1189 3 R10193 Insecticidal crystal 4.44e+02
39 41 64.1 1189 8 R39757 Delta endotoxin. 4.44e+02
40 41 64.1 1189 2 R10130 Lepidopteran-active t 4.44e+02
41 41 64.1 1190 14 R85150 CryIC/CryIA(B) altern 4.44e+02
42 40 62.5 489 29 W48242 Mouse MDM2. 5.51e+02
43 40 62.5 489 28 W42972 Amino acid sequence o 5.51e+02
44 40 62.5 489 28 W42997 Amino acid sequence o 5.51e+02
45 40 62.5 614 15 R82630 70K autoantigen, part 5.51e+02

ALIGNMENTS

RESULT 1
ID W19749 standard; Protein; 140 AA.
AC W19749;
DT 16-SEP-1997 (first entry)
DE Mouse inhibitor of apoptosis protein homologue MIHE.
KW Inhibitor of apoptosis protein; IAP; mammalian IAP homologue; MIHE;
KW degenerative disease; infectious disease; autoimmune disease;
KW cancer; therapy; diagnosis.
OS Mus musculus.
PN W09723501-A1.
PD 03-JUL-1997.
PF 20-DEC-1996; AU0827.
PR 22-DEC-1995; AU-007275.
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
PI Vaux DL;
DR WPI; 97-350966/32.
DR N-PSDB: T72714.
PT Isolated protein homologues of viral inhibitors of apoptosis - used
PT to modulate apoptosis for treatment of degenerative, infectious or
PT autoimmune diseases and cancer
CC MIHE (W19749) is a murine homologue of baculovirus inhibitor of
CC apoptosis protein (IAP). Its amino acid sequence was deduced from
CC an isolated nucleic acid (see also T72714) obtd. by a database
CC search for sequences homologous to a baculovirus IAP repeat (BIR)
CC consensus sequence (see also W19744). Unlike IAP, MIHE does not
CC contain a RING finger domain. IAP homologues (see also W19745-48
CC and W19750-52) and their derivatives and chemical analogues can be
CC used in methods for modulating apoptosis in animal cells,
CC specifically for treatment, by inhibition, of degenerative and
CC infectious disease or, by promotion, of cancer and autoimmune
CC disease.
SQ Sequence 140 AA;

Query Match 75.0%; Score 48; DB 23; Length 140;
Best Local Similarity 62.5%; Pred. No. 9.40e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 69 pddnpjee 76
Qy 3 PEDNPVED 10

RESULT 2
 ID R36534 standard; Protein; 928 AA.
 AC R36534; 1993 (first entry)
 DT 24-AUG-1993 (first entry)
 DE Retinoblastoma (RB) protein.
 KW RB gene product; p56RB portion; cell cycle progression control;
 KW combination; therapeutic methods; arrest; tumorigenesis;
 KW regulation; physiological processes; blood cell prodn.;
 KW gamete prodn.; ss.
 OS Homo sapiens SR-40 cell line.
 PN WO9308267-A.
 PD 29-APR-1993.
 PE 16-OCT-1992; U08918.
 PR 17-OCT-1991; US-778510.
 PA (REGC) UNIV CALIFORNIA.
 PI Goodrich DW, Lee EYP, Lee WH, Wang NP;
 DR WPI; 93-152462/18.
 DR N-PSDB; Q41545.
 PT Method of controlling cell cycle progression - uses purified
 PT retinoblastoma protein or fragment, for use in combination with
 PT therapeutic methods to arrest tumorigenesis
 PS Claim 10; Fig 9; 68pp; English.
 CC The sequence is that of the retinoblastoma (RB) protein which may be
 CC used as part of a method of controlling cell cycle progression which
 CC may be used in combination with therapeutic methods to arrest
 CC tumorigenesis in organisms. The cell cycle can be reversibly
 CC arrested in a convenient and safe manner. The protein is used in
 CC a compsn. is relatively inexpensive and readily obtainable, and
 CC shows little or no toxic effects on healthy cells. It is also
 CC compatible with other methods and devices for regulating certain
 CC physiological processes of the body, such as blood cell prodn. and
 CC gamete prodn. Fragments of the protein are soluble in low concns.
 CC of glycerol thus enhancing their value in pharmaceutical applns.
 CC Sequence 928 AA;
 SQ Sequence 928 AA;
 Query Match 70.3%; Score 45; DB 7; Length 928;
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 Db 27 pppepdpqd 36
 :|||:|:|:|
 QY 1 APPEDNPVED 10

RESULT 3
 ID R71680 standard; Protein; 928 AA.
 AC R71680;
 DT 16-OCT-1995 (first entry)
 DE Retinoblastoma p110RB protein.
 KW Rb110; retinoblastoma; cancer; tumor suppressor protein;
 KW cell cycle; p110RB.
 OS Homo sapiens.
 PN WO9507708-A.
 PD 23-MAR-1995.
 PE 13-SEP-1994; U10357.
 PR 13-SEP-1993; US-121108.
 PA (CANJ-) CANJI INC.
 PI Goodrich DW, Johnson D, Lee EYP, Lee W, Shepard HM;
 DR WPI; 95-131179/17.
 DR N-PSDB; Q86398.
 PT Admin. of a functional retinoblastoma polypeptide or protein -
 PT used to prevent and inhibit prim. and sec. retinoblastoma-linked
 PT cancers
 PS Disclosure; Fig.2; 163pp; English.
 CC By chromosomal walking from the esterase D gene on chromosome 13,
 CC the retinoblastoma susceptibility gene (RB) was identified on the
 CC basis of chromosomal location, homologous deletion and tumor-specific
 CC alterations in expression. Screening of cDNA libraries yielded clone
 CC RB-5 incorporating the complete RB cDNA sequence given in Q86398 and
 CC encoding a 110-140 kDa nuclear phosphoprotein named p110RB (R71680).
 CC Recombinant p110RB was prepared in E. coli, but with a P2A mutation
 CC for cloning convenience (R71681).
 CC Sequence 928 AA;
 SQ Sequence 928 AA;
 Query Match 70.3%; Score 45; DB 7; Length 928;
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 Db 27 pppepdpqd 36
 :|||:|:|:|
 QY 1 APPEDNPVED 10

RESULT 4
 ID R71681 standard; Protein; 928 AA.
 AC R71681;
 DT 16-OCT-1995 (first entry)
 DE Recombinant p110RB protein.
 KW Rb110; retinoblastoma; cancer; tumor suppressor protein;
 KW cell cycle; p110RB; Escherichia coli.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_difference 2
 FT /note= "amino acid at position 2 is Pro in the
 FT natural sequence"
 PN WO9507708-A.
 PD 23-MAR-1995.
 PE 13-SEP-1994; U10357.
 PR 13-SEP-1993; US-121108.
 PA (CANJ-) CANJI INC.
 PI (REGC) UNIV CALIFORNIA.
 PI Goodrich DW, Johnson D, Lee EYP, Lee W, Shepard HM;
 DR WPI; 95-131179/17.
 DR Admin. of a functional retinoblastoma polypeptide or protein -
 DR used to prevent and inhibit prim. and sec. retinoblastoma-linked
 DR cancers
 PS Disclosure; Fig.30; 163pp; English.
 CC By chromosomal walking from the esterase D gene on chromosome 13,
 CC the retinoblastoma susceptibility gene (RB) was identified on the
 CC basis of chromosomal location, homologous deletion and tumor-specific
 CC alterations in expression. Screening of cDNA libraries yielded clone
 CC RB-5 incorporating the complete RB cDNA sequence given in Q86398 and
 CC encoding a 110-140 kDa nuclear phosphoprotein named p110RB (R71680).
 CC Recombinant p110RB was prepared in E. coli, but with a P2A mutation
 CC for cloning convenience (R71681).
 CC Sequence 928 AA;
 SQ Sequence 928 AA;
 Query Match 70.3%; Score 45; DB 13; Length 928;
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 Db 27 pppepdpqd 36
 :|||:|:|:|
 QY 1 APPEDNPVED 10

RESULT 5
 ID W09411 standard; Protein; 928 AA.
 AC W09411;
 DT 22-JUN-1997 (first entry)
 DE Retinoblastoma susceptibility phosphoprotein pPRB110.
 KW Retinoblastoma susceptibility gene; pPRB110; nuclear phosphoprotein;
 KW cancer; osteosarcoma; fibrosarcoma; glioblastoma; breast cancer;
 KW polyclonal antibody; diagnosis.
 OS Homo sapiens.
 PN US5578701-A.
 PD 26-NOV-1996.
 PE 17-SEP-1987; 098612.
 PR 17-SEP-1987; US-098612.
 PR 11-JUL-1990; US-550877.
 PR 14-JUL-1992; US-914039.
 PR 17-JUN-1993; US-079207.

CC Use of recombinant p110RB will reduce the need for conventional
 CC radiotherapy or chemotherapy.
 SQ Sequence 928 AA;
 Query Match 70.3%; Score 45; DB 13; Length 928;
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 Db 27 pppepdpqd 36
 :|||:|:|:|
 QY 1 APPEDNPVED 10

RESULT 4
 ID R71681 standard; Protein; 928 AA.
 AC R71681;
 DT 16-OCT-1995 (first entry)
 DE Recombinant p110RB protein.
 KW Rb110; retinoblastoma; cancer; tumor suppressor protein;
 KW cell cycle; p110RB; Escherichia coli.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_difference 2
 FT /note= "amino acid at position 2 is Pro in the
 FT natural sequence"
 PN WO9507708-A.
 PD 23-MAR-1995.
 PE 13-SEP-1994; U10357.
 PR 13-SEP-1993; US-121108.
 PA (CANJ-) CANJI INC.
 PI (REGC) UNIV CALIFORNIA.
 PI Goodrich DW, Johnson D, Lee EYP, Lee W, Shepard HM;
 DR WPI; 95-131179/17.
 DR Admin. of a functional retinoblastoma polypeptide or protein -
 DR used to prevent and inhibit prim. and sec. retinoblastoma-linked
 DR cancers
 PS Disclosure; Fig.30; 163pp; English.
 CC By chromosomal walking from the esterase D gene on chromosome 13,
 CC the retinoblastoma susceptibility gene (RB) was identified on the
 CC basis of chromosomal location, homologous deletion and tumor-specific
 CC alterations in expression. Screening of cDNA libraries yielded clone
 CC RB-5 incorporating the complete RB cDNA sequence given in Q86398 and
 CC encoding a 110-140 kDa nuclear phosphoprotein named p110RB (R71680).
 CC Recombinant p110RB was prepared in E. coli, but with a P2A mutation
 CC for cloning convenience (R71681).
 CC Sequence 928 AA;
 SQ Sequence 928 AA;
 Query Match 70.3%; Score 45; DB 13; Length 928;
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 Db 27 pppepdpqd 36
 :|||:|:|:|
 QY 1 APPEDNPVED 10

RESULT 5
 ID W09411 standard; Protein; 928 AA.
 AC W09411;
 DT 22-JUN-1997 (first entry)
 DE Retinoblastoma susceptibility phosphoprotein pPRB110.
 KW Retinoblastoma susceptibility gene; pPRB110; nuclear phosphoprotein;
 KW cancer; osteosarcoma; fibrosarcoma; glioblastoma; breast cancer;
 KW polyclonal antibody; diagnosis.
 OS Homo sapiens.
 PN US5578701-A.
 PD 26-NOV-1996.
 PE 17-SEP-1987; 098612.
 PR 17-SEP-1987; US-098612.
 PR 11-JUL-1990; US-550877.
 PR 14-JUL-1992; US-914039.
 PR 17-JUN-1993; US-079207.

PR 08-APR-1994; US-225099.
 PA (REGC) UNIV CALIFORNIA.
 PI Lee EYP, Lee W;
 DR WPI; 97-020465/02.
 PT Retinoblastoma phosphoprotein pRB110-specific polyclonal antibody
 PT - for diagnosing retinoblastoma and other related tumours or
 PT susceptibility to them
 PS Disclosure; Fig 2; 21pp; English.
 CC A 110-114 kDa phosphoprotein product (W09411) of the retinoblastoma
 CC susceptibility gene is designated pRB110. Its sequence was deduced
 CC from a full-length cDNA sequence that included the sequence of a
 CC clone derived from the retinoblastoma cell line Y79. pRB110 is
 CC primarily located in the cell nucleus and has DNA binding activity.
 CC The absence or loss of pRB10 mediates oncogenicity. pRB110-
 CC specific polyclonal antibodies can be utilised in methods of
 CC diagnosing hereditary predisposition to retinoblastoma or to other
 CC diseases controlled by the retinoblastoma gene such as
 CC osteosarcoma, fibrosarcoma, glioblastoma and breast cancer.
 CC Provision of pRB110 to an individual through molecular induction
 CC and gene transplanting may be used as a means of suppressing
 CC tumorigenesis.
 SQ Sequence 928 AA;

Query Match 70.3%; Score 45; DB 22; Length 928;
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 ppspeeqd 36
 :||||:|:
 QY 1 APPEDNPVED 10

RESULT 6
 ID R74271 standard; Protein; 928 AA.

AC R74271;
 DT 28-NOV-1995 (first entry)
 DE Retinoblastoma tumour suppressor protein.
 KW Recombinant; adenovirus; expression vector; TSG; small lung cancer;
 KW hepatocarcinoma; melanoma; retinoblastoma; sarcoma; sickle cell;
 KW anaemia; Tay-Sach's disease.
 OS Homo sapiens.
 PN W09511984-A.
 PD 04-MAY-1995.
 PE 25-OCT-1994; U12235.
 PR 25-OCT-1993; US-142669.
 PR 19-MAY-1994; US-246007.
 PA (CANU-) CANJI INC.
 PI Gregory RJ, Maneval DC, Wills KN;
 DR WPI; 95-178876/23.
 DR N-PSDB; Q90059.
 PT Adenoviral vector with deletion of viral protein IX contains
 PT foreign gene - esp. encoding tumour suppressor protein for gene
 PT therapy of tumours, reduces contamination by wild type adenovirus
 PS Disclosure; Fig 3; 92pp; English.
 CC The sequence is that of a retinoblastoma tumour protein. The gene
 CC encoding this protein may be used in a novel method involving a
 CC recombinant adenovirus expression vector to treat diseases
 CC associated with the absence of the TSG or the presence of a mutated
 CC TSG, e.g. many forms of carcinoma, sickle cell anaemia or Tay-Sach's
 CC disease.
 CC See also R74272.
 SQ Sequence 928 AA;

Query Match 70.3%; Score 45; DB 14; Length 928;
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 ppspeeqd 36
 :||||:|:
 QY 1 APPEDNPVED 10

RESULT 7

ID R05305 standard; Protein; 928 AA.
 AC R05305;
 DT 11-OCT-1990 (first entry)
 DE Cancer suppressing gene (CSG) product.
 KW Cancer; cancer suppressing gene; CSG; 13q14; retinoblastoma;
 KW RB; ds.
 OS Homo sapiens.
 PN W09005180-A.
 PD 17-MAY-1990.
 PE 30-OCT-1989; 004808.
 PR 31-OCT-1988; US-265829.
 PA (REGC) Univ of California.
 PI Lee WH, Huang HJS;
 DR WPI; 90-178822/23.
 DR N-PSDB; Q04713.
 PT Controlling cancer -
 PT by replacing ineffective cancer suppressing gene with cloned,
 PT active gene.
 PS Claim 35; Page 86; 105pp; English.
 CC Gene is taken from human chromosome 13q14 retinoblastoma (RB) cDNA.
 CC By installing a working CSG, safe and specific treatment and
 CC prophylaxis can be given to cancer patients.
 SQ Sequence 928 AA;

Query Match 70.3%; Score 45; DB 1; Length 928;
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 ppspeeqd 36
 :||||:|:
 QY 1 APPEDNPVED 10

RESULT 8

ID R06289 standard; Protein; 928 AA.
 AC R06289;
 DT 13-DEC-1990 (first entry)
 DE Predicted retinoblastoma gene product.
 KW Osteosarcoma; fibrosarcoma; glioblastoma; breast cancer; ds.
 OS Homo sapiens.
 PN U54942123-A.
 PD 17-JUL-1990.
 PE 17-SEP-1987; 098612.
 PR 17-SEP-1987; US-098612.
 PA (REGC) UNIV OF CALIFORNIA.
 PI Lee WH, Eva Y, Lee HP;
 DR WPI; 90-245977/32.
 PT Diagnosing absence or inactivation of retinoblastoma gene - by
 PT detecting the absence of specific anti-pRB 110 antibody
 PT immuno-complex formed using tissue
 PS Disclosure; P; English.
 CC Labelled Abs raised to the RB gene product may be used to screen
 CC for RB and in diagnosis of susceptibility to associated secondary
 CC cancers such as osteosarcoma, fibrosarcoma, glioblastoma and
 CC breast cancer.
 SQ Sequence 928 AA;

Query Match 70.3%; Score 45; DB 1; Length 928;
 Best Local Similarity 50.0%; Pred. No. 1.84e+02;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 ppspeeqd 36
 :||||:|:
 QY 1 APPEDNPVED 10

RESULT 9

ID R59922 standard; Protein; 2485 AA.
 AC R59922;
 DT 22-FEB-1995 (first entry)
 DE RAS associated GAP NF204.
 KW Ras; GTPase activating protein; GAP; GAP related domain; GRD;
 KW pK10; pK11; Saccharomyces cerevisiae; RAS2; v-Ras; heat shock;

KW neurofibromatosis type 1; NF1.
 OS Homo sapiens.
 PN W09416069-A.
 PD 21-JUL-1994.
 PF 12-JAN-1994; U00198.
 PR 15-JAN-1993; US-004824.
 PA (SCHE) SCHERING CORP.
 PI Kaziro Y, Nakafuku M;
 DR WPI: 94-249216/30.
 PT Blocking Ras-induced effects on a cell - by introducing a GTPase
 PT activating protein to the cell, used esp. in treatment of cancers
 PS Disclosure; Page 44-52; 87pp; English.
 CC Human neurofibromatosis type 1 (NF1)-GAP related domain (GRD) was
 CC cloned into the yeast expression vector pK10 to obtain pKp11. The
 CC pKp11 DNA was mutagenized by hydroxylamine in vitro and transformed
 CC into S. cerevisiae TK161-R2V-D, which carries an oncogenic-type
 CC RAS2Val19 mutation. The heat shock sensitivity of the clones was
 CC checked. Plasmid DNAs were recovered, re-transformed into TK161-
 CC R2V-D, and phenotypic reversion was examined. 2 Clones, NF201
 CC (given in R59921) and NF204, which had strong suppression activity
 CC for RAS2Val19, were selected. The mutant NF1-GRDs were also able
 CC to inhibit v-Ras-induced transformation in mammalian cells.
 SQ Sequence 2485 AA;

Query Match 70.3%; Score 45; DB 11; Length 2485;
 Best Local Similarity 60.0%; Pred. No. 1.84e+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1192 gpehpkpvd 1201
 :||| :|||
 QY 1 APPEDNPVED 10

RESULT 10
 ID R59921 standard; protein; 2485 AA.
 AC R59921;
 DT 22-FEB-1995 (first entry)
 DE RAS associated GAP NF201.
 KW Ras; GTPase activating protein; GAP; GAP related domain; GRD;
 KW pK10; pKp11; Saccharomyces cerevisiae; RAS2; v-Ras; heat shock;
 KW neurofibromatosis type 1; NF1.
 OS Homo sapiens.
 PN W09416069-A.
 PD 21-JUL-1994.
 PF 12-JAN-1994; U00198.
 PR 15-JAN-1993; US-004824.
 PA (SCHE) SCHERING CORP.
 PI Kaziro Y, Nakafuku M;
 DR WPI: 94-249216/30.
 PT Blocking Ras-induced effects on a cell - by introducing a GTPase
 PT activating protein to the cell, used esp. in treatment of cancers
 PS Disclosure; Page 36-44; 87pp; English.
 CC Human neurofibromatosis type 1 (NF1)-GAP related domain (GRD) was
 CC cloned into the yeast expression vector pK10 to obtain pKp11. The
 CC pKp11 DNA was mutagenized by hydroxylamine in vitro and transformed
 CC into S. cerevisiae TK161-R2V-D, which carries an oncogenic-type
 CC RAS2Val19 mutation. The heat shock sensitivity of the clones was
 CC checked. Plasmid DNAs were recovered, re-transformed into TK161-
 CC R2V-D, and phenotypic reversion was examined. 2 Clones, NF201 and
 CC NF204 (given in R59922), which had strong suppression activity for
 CC RAS2Val19, were selected. The mutant NF1-GRDs were also able to
 CC inhibit v-Ras-induced transformation in mammalian cells.
 SQ Sequence 2485 AA;

Query Match 70.3%; Score 45; DB 11; Length 2485;
 Best Local Similarity 60.0%; Pred. No. 1.84e+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1192 gpehpkpvd 1201
 :||| :|||
 QY 1 APPEDNPVED 10

RESULT 11
 ID W13280 standard; Protein; 2818 AA.
 AC W13280;
 DT 05-JUN-1997 (first entry)
 DE Human neurofibromin.
 KW Human; neurofibromin; yeast; IRA; protein; inhibition; GTPase;
 KW regulation; ras-CAMP; pathway; mammalian; GAP; ras p21; gene;
 KW activation; neurofibromatosis; type 1; NF1; somatic; mutation;
 KW tumour; detection; diagnosis; prognosis; defective; treatment.
 OS Homo sapiens.
 FH Key
 FT Location/Qualifiers
 FT domain 1175..1534
 FT /note= "GTPase activating protein (GAP) related
 FT domain (GRD)."
 FT region 1389..1391
 FT /note= "conserved region in GRD"
 FT US5605799-A.
 PN 25-FEB-1997.
 PD 12-JUL-1990; 551531.
 PF 12-JUL-1990; US-551531.
 PR 16-APR-1993; US-047088.
 PR 28-MAR-1995; US-411389.
 PA (UTAH) UNIV UTAH RES FOUND.
 PI Cawthon RM, Li Y, White RL;
 DR WPI: 97-153572/14.
 DR N-PSDB; T45941.
 PT Detection of defective ras regulation at the neurofibromatosis type
 PT 1 gene in tumour - by detecting mutation in specified region of gene
 PS Claim 1; Columns 17-38; 35pp; English.
 CC The present sequence is human neurofibromin (hNF), which is
 CC largely homologous to yeast IRA protein (inhibitory regulators of
 CC the ras-CAMP pathway) and mammalian GAP (ras p21 GTPase activating
 CC proteins). The hNF gene is the human neurofibromatosis type 1 (NF1)
 CC gene, somatic mutations of which in the region spanning nucleotides
 CC 3809-4888 of the NF1 cDNA, in human tumours, indicates defective
 CC ras regulation. Therefore a tumour found to contain a somatic
 CC mutation in the NF1 gene can be treated using ras activity as the
 CC focus, whereas a tumour not containing such a mutation will require
 CC other courses of treatment. A tumour containing a somatic mutation
 CC in the NF1 gene can be treated by inactivating ras p21, also as GAP
 CC p120 is present, but apparently latent, inhibition of GDP/GTP exchange would
 CC be beneficial and finally inhibition of GDP/GTP exchange would also
 CC counteract the loss of hNF or hNF GAP related domain activity.
 CC Sequence. 2818 AA;
 SQ

Query Match 70.3%; Score 45; DB 21; Length 2818;
 Best Local Similarity 60.0%; Pred. No. 1.84e+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1526 gpehpkpvd 1535
 :||| :|||
 QY 1 APPEDNPVED 10

RESULT 12
 ID R22268 standard; Protein; 2818 AA.
 AC R22268;
 DT 06-MAY-1992 (first entry)
 DE NF1 gene product.
 KW von Recklinghausen neurofibromatosis disease; autosomal dominant;
 KW gene therapy.
 OS Homo sapiens.
 PN W09200387-A.
 PD 09-JAN-1992.
 PF 28-JUN-1991; U04624.
 PR 29-JUN-1990; US-547090.
 PA (UNMI) UNIV OF MICHIGAN.
 PI Collins FS, Wallace MR, Marchuk DA, Andersen LB, Gutmann DH;
 DR WPI: 92-041568/05.
 DR N-PSDB; Q20602.
 PT DNA sequences to von-Recklinghausen neurofibromatosis gene - and
 PT derived amino acid sequences and probes for screening NF1 in early
 PT stages of disease

PS Claim 25; Page 67; 122pp; English.
 CC This is the amino acid sequence of the von Recklinghausen neuro-
 CC fibromatosis (Nf1) gene product. It and antibodies raised to it
 CC can be used in hybridisation and immunological assays to screen for
 CC the presence of a normal or defective Nf1 gene product. Functional
 CC assays to measure levels of gene function can also be used for
 CC diagnosis or to monitor treatment. Patient therapy through
 CC supplementation with the normal Nf1 product which can be
 CC produced by recombinant techniques is also possible.
 SQ Sequence 2818 AA;

Query Match 70.3%; Score 45; DB 4; Length 2818;
 Best Local Similarity 60.0%; Pred. No. 1.84e+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1526 gpehkpvd 1535
 :|||:||||
 QY 1 APPEDNPVED 10

RESULT 13
 ID R86553 standard; Protein; 395 AA.
 AC R86553;
 DT 31-MAR-1996 (first entry)
 DE Vibrio cholerae strain 395 accessory cholera enterotoxin ACE protein.
 KW Bacterium; enteropathogen; cholera enterotoxin; choleragen;
 KW cholera toxin; diarrhea; vaccine; plasmid pCV620; ACE.
 OS Vibrio cholerae.
 PN US5470729-A.
 PD 28-NOV-1995.
 PF 04-MAR-1983; 472276.
 PR 04-MAR-1983; US-472276.
 PR 17-FEB-1984; US-581406.
 PR 27-MAY-1986; US-867633.
 PR 05-JUN-1989; US-363383.
 PR 05-JUN-1990; US-533315.
 PR 16-JAN-1992; US-821872.
 PR 12-AUG-1992; US-931943.
 PA (UYMA-) UNIV MARYLAND BALTIMORE.
 PI Baudry-maurelli B, Fasano A, Kaper JB;
 DR WPI; 96-019870/02.
 DR N-PSDB; T06601.
 DT New avirulent Vibrio cholerae strains - comprise deletions in the
 DT cholera toxin and zonula occludens toxin genes, for vaccination
 PT against cholera
 PT Example 12; Column 29-36; 55pp; English.
 CC This is the 11 kDa protein sequence of accessory cholera
 CC enterotoxin encoded by a 297 bp ORF (accessory cholera enterotoxin;
 CC ACE) in plasmid pCV620. This protein may be useful in the
 CC construction of cholera vaccines.
 SQ Sequence 395 AA;

Query Match 68.8%; Score 44; DB 16; Length 395;
 Best Local Similarity 55.6%; Pred. No. 2.30e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 121 sppsdspvd 129
 :|||:||||
 QY 1 APPEDNPVE 9

RESULT 14
 ID R86555 standard; Protein; 395 AA.
 AC R86555;
 DT 31-MAR-1996 (first entry)
 DE Vibrio cholerae El Tor accessory cholera enterotoxin ACE protein.
 KW Bacterium; enteropathogen; cholera enterotoxin; choleragen;
 KW cholera toxin; diarrhea; vaccine; plasmid pCV620; ACE.
 OS Vibrio cholerae.
 PN US5470729-A.
 PD 28-NOV-1995.
 PF 04-MAR-1983; 472276.
 PR 04-MAR-1983; US-472276.

PR 17-FEB-1984; US-581406.
 PR 27-MAY-1986; US-867633.
 PR 05-JUN-1989; US-363383.
 PR 05-JUN-1990; US-533315.
 PR 16-JAN-1992; US-821872.
 PR 12-AUG-1992; US-931943.
 PA (UYMA-) UNIV MARYLAND BALTIMORE.
 PI Baudry-maurelli B, Fasano A, Kaper JB;
 DR WPI; 96-019870/02.
 DR N-PSDB; T06602.

PT New avirulent Vibrio cholerae strains - comprise deletions in the
 PT cholera toxin and zonula occludens toxin genes, for vaccination
 PT against cholera
 PS Example 12; Column 37-44; 55pp; English.
 CC This is the 11 kDa protein sequence of accessory cholera
 CC enterotoxin encoded by a 297 bp ORF (accessory cholera enterotoxin;
 CC ACE) in plasmid pCV620. This protein may be useful in the
 CC construction of cholera vaccines.
 SQ Sequence 395 AA;

Query Match 68.8%; Score 44; DB 16; Length 395;
 Best Local Similarity 55.6%; Pred. No. 2.30e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 121 sppsdspvd 129
 :|||:||||
 QY 1 APPEDNPVE 9

RESULT 15
 ID R08390 standard; protein; 437 AA.
 AC R08390;
 DT 25-FEB-1991 (first entry)
 DE Truncated form of human FR-X protein.
 KW MHC class II; down regulation; autoimmune disease; HLA promoter.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT domain 255..376
 FT W09012812-A.
 PN 01-NOV-1990.
 PF 18-APR-1990; E00625.
 PR 18-APR-1989; EP-106944.
 PR 14-AUG-1989; EP-115008.
 PA (MACH/) Mach B.
 PI Mach B;
 DT WPI; 90-348429/46.
 PT Purified proteins and compans. - regulate expression of MHC class II
 PT genes and bind to controlling DNA sequences.
 PS Claim 6; Fig 10; 83pp; English.
 CC The protein which regulates the expression of MHC class II genes by
 CC binding to DNA sequences which control this expression. It causes
 CC down-regulation, useful for the prevention and treatment of auto-
 CC immune diseases such as Insulin Dependent diabetes, Multiple
 CC Sclerosis, Lupus Erythematosus and Rheumatoid Arthritis. It can
 CC also be used for screening and identifying substances capable of
 CC inhibiting the expression of the MHC II genes. A polypeptide
 CC comprising only the DNA binding domain (see feature table) will
 CC bind to the X box of the HLA class II promoter.
 CC See also R08338, R07661 and R08391.
 SQ Sequence 437 AA;

Query Match 68.8%; Score 44; DB 2; Length 437;
 Best Local Similarity 55.6%; Pred. No. 2.30e+02;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 53 sppsdspvq 61
 :|||:||||
 QY 1 APPEDNPVE 9

Search completed: Thu May 13 15:27:52 1999
 Job time : 17 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 13 15:29:15 1999; MasPar time 4.04 Seconds

Tabular output not generated. 123.105 Million cell updates/sec

Title: >US-09-040-485-6
Description: (1-10) from US09040485.pep
Perfect Score: 64
Sequence: 1 APPENDNPVED 10

Scoring table: PAM 150
Gap 15

Searched: 165420 seqs, 49795644 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sprenbl6

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 21.054; Variance 25.580; scale 0.823

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	49	76.6	225	11	APC-BINDING PROTEIN EB	2.59e+00
2	48	75.0	140	11	TIAP.	4.27e+00
3	48	75.0	1175	4	NPAT.	4.27e+00
4	48	75.0	1427	4	Q14207	4.27e+00
5	47	73.4	129	2	E14 PROTEIN.	6.99e+00
6	46	71.9	142	4	HYPOTHETICAL 13.0 KD P	1.14e+01
7	46	71.9	142	4	APOPTOSIS INHIBITOR SU	1.14e+01
8	46	71.9	146	10	GLUTATHIONE S-TRANSFER	1.14e+01
9	45	70.3	2042	5	3D7VAR1 (FRAGMENT).	1.14e+01
10	45	70.3	53	4	RETINOBLASTOMA SUSCEPT	1.83e+01
11	45	70.3	86	2	HYPOTHETICAL 9.9 KD PR	1.83e+01
12	45	70.3	97	5	DNA FRAGMENT WITH HIGH	1.83e+01
13	45	70.3	258	2	DEHYDROGENASE.	1.83e+01
14	45	70.3	287	5	CO3H5_3 PROTEIN.	1.83e+01
15	45	70.3	317	4	RNA POLYMERASE III SUB	1.83e+01
16	45	70.3	339	2	JUGLANDIS COPPER-RESIS	1.83e+01
17	45	70.3	378	4	NEUROFIBROMATOSIS PROT	1.83e+01
18	45	70.3	840	4	NEUROFIBROMATOSIS 1 (F	1.83e+01
19	45	70.3	1294	4	KIAA0293 (FRAGMENT).	1.83e+01
20	45	70.3	1426	11	CUT-LIKE 2 (CUX-2).	1.83e+01
			2172	5	HYPOTHETICAL 236.2 KD	1.83e+01
			Q09515			

21	45	70.3	2820	11	P97526	1.83e+01
22	44	68.8	109	5	O02443	2.94e+01
23	44	68.8	122	2	O68696	2.94e+01
24	44	68.8	131	1	Q49612	2.94e+01
25	44	68.8	259	1	O29756	2.94e+01
26	44	68.8	273	13	Q91257	2.94e+01
27	44	68.8	304	2	Q60225	2.94e+01
28	44	68.8	334	10	O24272	2.94e+01
29	44	68.8	348	13	O57378	2.94e+01
30	44	68.8	365	4	O60509	2.94e+01
31	44	68.8	473	5	Q26236	2.94e+01
32	44	68.8	598	2	Q50866	2.94e+01
33	44	68.8	739	2	Q59557	2.94e+01
34	44	68.8	740	2	O05763	2.94e+01
35	44	68.8	1262	5	Q20684	2.94e+01
36	44	68.8	1506	13	P79927	2.94e+01
37	44	68.8	2764	5	O01399	2.94e+01
38	44	68.8	2802	5	O01398	2.94e+01
39	44	68.8	2802	5	O01397	2.94e+01
40	44	68.8	3413	2	O54593	2.94e+01
41	43	67.2	149	14	P90432	4.68e+01
42	43	67.2	506	14	O57058	4.68e+01
43	43	67.2	506	14	O41911	4.68e+01
44	43	67.2	507	14	P89153	4.68e+01
45	43	67.2	1103	14	O41928	4.68e+01

ALIGNMENTS

RESULT 1
ID Q61167; PRELIMINARY; PRT; 225 AA.
AC Q61167;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMREL. 01, LAST ANNOTATION UPDATE)
DE APC-BINDING PROTEIN EB2 (FRAGMENT).
GN EB2.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RA HOFFMAN N.G., KAY B.K.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U51204; G1256436; -;
FT NON_TER 1 1
SQ SEQUENCE 225 AA; 25678 MW; 812588B4 CRC32;

Query Match 76.6%; Score 49; DB 11; Length 225;
Best Local Similarity 60.0%; Pred. No. 2.59e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 207 APPEDDEIEE 216

|||||: -;
Qy 1 APPEDNPVED 10

RESULT 2

ID O70201; PRELIMINARY; PRT; 140 AA.
AC O70201;
DT 01-AUG-1998 (TREMREL. 07, CREATED)
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE TIAP.
GN TIAP.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-EMBRYO;
RA KOBAYASHI K., OTAKI M., OGASAWARA T., TOKUHISA T.;
SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL; AB013819; D1029206; -
SQ SEQUENCE 140 AA; 16297 MW; 2BD48871 CRC32;

Query Match 75.0%; Score 48; DB 11; Length 140;
Best Local Similarity 62.5%; Pred. No. 4.27e+00;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 69 PDDNPIEE 76
I:||||:;
Qy 3 PEDNPVED 10

RESULT 3
ID Q13632 PRELIMINARY; PRT; 1175 AA.

AC Q13632;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE NPAT.
GN NPAT.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]

RP SEQUENCE FROM N.A.
RA CHEN X., YANG L., UDAR N., LIANG T., XU S., UHRHAMMER N., BAY J.O.,
RA WANG Z., DANDAKAR U., CHIPLONKAR S., KLISAK I., TELATAR M., YANG H.,
RA CANNON P., GATTI R.A.,
RA MAMM. GENOME 0:0-0(0).
RL EMBL; U58852; G1381667; -
SQ SEQUENCE 1175 AA; 126731 MW; CFC49BD6 CRC32;

Query Match 75.0%; Score 48; DB 4; Length 1175;
Best Local Similarity 60.0%; Pred. No. 4.27e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 688 TPPECTPVEN 697
I:||||:
Qy 1 APPEDNPVED 10

RESULT 4
ID Q14207 PRELIMINARY; PRT; 1427 AA.

AC Q14207; Q14967;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)
DE E14 PROTEIN.
GN E14.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RA COOPER P.R., BYRD P.J., TAYLOR A.M.R.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]

RP SEQUENCE FROM N.A.
RA BYRD P.J., MCCONVILLE C.M., COOPER P.R., PARKHILL J., MCGUIRE G.,
RA STANKOVIC T., THICK J., TAYLOR A.M.R.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-PLACENTA, AND TESTIS;
RX MEDLINE; 96338579.
RA IMAI T., YAMAUCHI M., SEKI N., SUGAWARA T., SAITO T., MATSUDA Y.,
RA ITOH H., NAGASE T., NOMURA N., HORI T.;
RL GENOME RES. 6:439-447(1996).
DR EMBL; X97186; E238806; -
DR EMBL; D83243; G1304114; -

FT CONFLICT 14 14 Y -> N (IN REF. 2).
FT CONFLICT 295 295 L -> I (IN REF. 3).
FT CONFLICT 471 471 N -> Y (IN REF. 3).

FT CONFLICT 967 967 E -> Q (IN REF. 3).
FT CONFLICT 973 973 V -> L (IN REF. 3).
FT CONFLICT 987 987 A -> V (IN REF. 3).
SQ SEQUENCE 1427 AA; 154273 MW; 2DDEBF4 CRC32;

Query Match 75.0%; Score 48; DB 4; Length 1427;
Best Local Similarity 60.0%; Pred. No. 4.27e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 688 TPPECTPVEN 697
I:||||:
Qy 1 APPEDNPVED 10

RESULT 5
ID O50383 PRELIMINARY; PRT; 129 AA.

AC O50383;
DT 01-JUN-1998 (TREMREL. 06, CREATED)
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 13.0 KD PROTEIN.
GN MT004.11.
OS MYCOBACTERIUM TUBERCULOSIS.
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA OLIVER K., SKELTON J., BADCOCK K., CHURCHER C.M., HARRIS D.;
RL SUBMITTED (DEC-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE; 96181548.
RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
RA COLE S.T.;
RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).
DR EMBL; AL009198; E1202271; -
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 129 AA; 12988 MW; 17B14A83 CRC32;

Query Match 73.4%; Score 47; DB 2; Length 129;
Best Local Similarity 60.0%; Pred. No. 6.99e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 28 APAQANPVDD 37
I:||||:
Qy 1 APPEDNPVED 10

RESULT 6
ID O15392 PRELIMINARY; PRT; 142 AA.

AC O15392;
DT 01-JAN-1998 (TREMREL. 05, CREATED)
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE APOPTOSIS INHIBITOR SURVIVIN.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 941148797.
RA ALTIERI D.C.;
RL J. BIOL. CHEM. 269:3139-3142(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95034823.

RA ALTIERI D.C.;
RL BIOCHEMISTRY 33:13848-13855(1994).
RN [3]

RP SEQUENCE FROM N.A.
RA AMBROSINI G., ADIDA C., ALTIERI D.C.;
RL NAT. MED. 3:917-921(1997).

DR EMBL: U75285; G2315863; -.

DR PFAM: PF00653; BIR.

SQ SEQUENCE 142 AA; 16389 MW; 702152A2 CRC32;

Query Match 71.9%; Score 46; DB 4; Length 142;
Best Local Similarity 50.0%; Pred. No. 1.14e+01;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 69 PDDPIEE 76
QY 3 PEDNPVED 10

RESULT 7
ID O65858 PRELIMINARY; PRT; 146 AA.

AC O65858;

DT 01-AUG-1998 (TREMBLREL. 07, CREATED)

DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)

DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE GLUTATHIONE S-TRANSFERASE (FRAGMENT).

GN RGST II.

OS ORYZA SATIVA (RICE).

OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;

CC CYPERALES; GRAMINEAE.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CV. LEMONT;

RA WU J., CRAMER C.L., HATZIOS K.K.;

RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL: AJ002381; E1289811; -.

KW TRANSFERASE.

FT NON_TER 146 146

SQ SEQUENCE 146 AA; 16443 MW; AC010F6D CRC32;

Query Match 71.9%; Score 46; DB 10; Length 146;
Best Local Similarity 50.0%; Pred. No. 1.14e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 77 APDEKVEE 86
QY 1 APPEDNPVED 10

RESULT 8

ID Q25766 PRELIMINARY; PRT; 2042 AA.

AC Q25766;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)

DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)

DE 3D7VAR1 (FRAGMENT).

GN 3D7VAR1.

OS PLASMODIUM FALCIPARUM.

OC EUKARYOTA; PROTOZOA; APICOMPLEXA; SPOROZOA; COCCIDIA; EUCCOCIDIIDA.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 96324414.

RA RUBIO J.P., THOMPSON J.K., COWMAN A.F.;

RL EMBO J. 15:4069-4077(1996).

DR EMBL: U53324; G1297091; -.

FT NON_TER 1 1

SQ SEQUENCE 2042 AA; 233669 MW; 77999777 CRC32;

Query Match 71.9%; Score 46; DB 5; Length 2042;
Best Local Similarity 100.0%; Pred. No. 1.14e+01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 658 PEDNPVE 664
QY 3 PEDNPVE 9

RESULT 9

AC Q92728 PRELIMINARY; PRT; 53 AA.

DT 01-FEB-1997 (TREMBLREL. 02, CREATED)

DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)

DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)

DE RETINOBLASTOMA SUSCEPTIBILITY PROTEIN (RB).

GN RB1.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

CC EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 88320373.

RA LEE E.Y., BOOKSTEIN R., YOUNG L.J., LIN C.J., ROSENFELD M.G.;

RA LEE W.H.;

RL PROC. NATL. ACAD. SCI. U.S.A. 85:6017-6021(1988).

DR EMBL: M19701; G190968; -.

SQ SEQUENCE 53 AA; 5392 MW; 2D5C3E39 CRC32;

Query Match 70.3%; Score 45; DB 4; Length 53;
Best Local Similarity 50.0%; Pred. No. 1.83e+01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 PPDEEDPEQD 36
QY 1 APPEDNPVED 10

RESULT 10

ID P74580 PRELIMINARY; PRT; 86 AA.

AC P74580;

DT 01-FEB-1997 (TREMBLREL. 02, CREATED)

DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)

DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE HYPOTHETICAL 9.9 KD PROTEIN.

OS SYNECHOCYSTIS SP.

OC EUBACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-PCC6803;

RA TABATA S.;

RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-PCC6803;

RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.;

RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.;

RA HOSOUCHI T., MATSUNO A., MORAKI A., NAKAZAKI N., NARUO K., OKUMURA S.;

RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.;

RA TABATA S.;

RL DNA RES. 3:109-136(1996).

DR EMBL: D90916; G1653776; -.

KW HYPOTHETICAL PROTEIN.

SQ SEQUENCE 86 AA; 9909 MW; B95C9013 CRC32;

Query Match 70.3%; Score 45; DB 2; Length 86;
Best Local Similarity 62.5%; Pred. No. 1.83e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 49 PDDTPVEE 56
QY 3 PEDNPVED 10

RESULT 11

ID Q26168 PRELIMINARY; PRT; 97 AA.

AC Q26168;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE DNA FRAGMENT WITH HIGHLY REPEATED SEQUENCE (FRAGMENT)
 OS PLASMODIUM VIVAX
 OC EUKARYOTA; PROTOZOA; APICOMPLEXA; SPOROZOA; COCCIDIA; EUCCOCCIDIIDA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA CAMPBELL J.R., FRANK E.D.;
 RL SUBMITTED (APR-1989) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; X15129; G10087; -
 FT NON_TER 1
 SQ SEQUENCE 97 AA; 9736 MW; 3866D3C5 CRC32;

Query Match 70.3%; Score 45; DB 5; Length 97;
 Best Local Similarity 71.4%; Pred. No. 1.83e+01;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 13 PPEENPI 19
 |||:|:|:
 QY 2 PPDNPV 8

RESULT 12
 ID O33339 PRELIMINARY; PRT; 258 AA.
 AC O33339;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE DEHYDROGENASE.
 GN MTV003.03C.
 OS MYCOBACTERIUM TUBERCULOSIS.
 OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RA MURPHY L., HARRIS D.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE; 96181548.
 RA PHILIPP W.J., POULET S., EIGLMEIER K., PASCOPELLA L.,
 RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,
 RA COLE S.T.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 93:3132-3137(1996).
 DR EMBL; AL008883; E1172949; -
 DR PFAM; PF00106; adh_short.
 DR PFAM; PF00678; adh_short.C2.
 SQ SEQUENCE 258 AA; 26804 MW; 60151FCC CRC32;

Query Match 70.3%; Score 45; DB 2; Length 258;
 Best Local Similarity 60.0%; Pred. No. 1.83e+01;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 93 SPEDNLIEN 102
 :|||||:|:
 QY 1 APPEDNPVED 10

RESULT 13
 ID O16657 PRELIMINARY; PRT; 287 AA.
 AC O16657;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE C03H5.3 PROTEIN.
 GN C03H5.3.

OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE; 94150718.
 RA BURTON J., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
 RA WILSON R., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
 RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
 RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
 RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
 RA MURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
 RA KIFURU L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
 RA SONNHAMMER E., STADEN R., SULSTON J., THERRY-MIEG J., THOMAS K.,
 RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
 RA WILKINSON-SPROAT J., WOHLDMANN P.;
 RL NATURE 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA ROHLFING T., WOHLDMANN P.;
 RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AF016674; G2315689; -
 SQ SEQUENCE 287 AA; 31998 MW; 4D4ED5ED CRC32;

Query Match 70.3%; Score 45; DB 5; Length 287;
 Best Local Similarity 50.0%; Pred. No. 1.83e+01;
 Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 80 APSOENSVD 89
 |||:|:|:|:
 QY 1 APPEDNPVED 10

RESULT 14
 ID O15319 PRELIMINARY; PRT; 317 AA.
 AC O15319;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE RNA POLYMERASE III SUBUNIT.
 GN RPC39.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 97315201.
 RA WANG Z., ROEDER R.G.;
 RL GENES DEV. 11:1315-1326(1997).
 DR EMBL; U93869; G2228752; -
 SQ SEQUENCE 317 AA; 36144 MW; BCC34DCF CRC32;

Query Match 70.3%; Score 45; DB 4; Length 317;
 Best Local Similarity 62.5%; Pred. No. 1.83e+01;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 10 PPDADPVE 17
 |||:|:|:
 QY 2 PPDNPVE 9

RESULT 15
 ID Q56796 PRELIMINARY; PRT; 339 AA.
 AC Q56796;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)

DE JUGLANDIS COPPER-RESISTANCE GENES, COMPLETE CDS.
OS XANTHOMONAS CAMPESTRIS.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC PSEUDOMONADACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COPPER-RESISTANCE;
RX MEDLINE; 94110224.
RA LEE Y.A., HENDSON M., PANOPoulos N.J., SCHROTH M.N.;
RL J. BACTERIOL. 176:173-188(1994).
DR EMBL; L19222; G461142; -.
SQ SEQUENCE 339 AA; 37041 MW; D1400210 CRC32;

Query Match 70.3%; Score 45; DB 2; Length 339;
Best Local Similarity 60.0%; Pred. NO. 1.83e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 121 PPSDHPVHD 130
QY 1 APPENDVED 10

Search completed: Thu May 13 15:29:42 1999
Job time : 27 secs.

W A I T E S R E L A

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 13 15:31:38 1999; Maspar time 2.52 Seconds

Tabular output not generated. 64.113 Million cell updates/sec

Title: >US-09-040-485-7
Description: (1-10) from US09040485.pep
Perfect Score: 66
Sequence: 1 BEQEVPPDT 10

Scoring table: PAM 150
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29

Statistics: Mean 14.792; Variance 49.512; scale 0.299

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	45	68.2	3144	29	W44742 Human huntingtin prot	1.72e+02
2	45	68.2	3144	11	R58777 Protein encoded by Hu	1.72e+02
3	45	68.2	3144	26	W36887 Previously undescribe	1.72e+02
4	45	68.2	3144	22	W09871 Human huntingtin	1.72e+02
5	43	65.2	140	1	R05669 Gamma-chicken atrial	2.72e+02
6	43	65.2	414	16	R30355 Transforming growth f	2.72e+02
7	43	65.2	414	14	R73597 Human TGF-beta 2 prot	2.72e+02
8	43	65.2	442	1	P91899 Sequence encoded by h	2.72e+02
9	43	65.2	442	16	R79922 Human transforming gr	2.72e+02
10	43	65.2	442	4	R20125 Sequence of human tra	2.72e+02
11	43	65.2	442	1	R05718 Human TGF-Beta2-442 p	2.72e+02
12	43	65.2	700	1	R03663 Human myb related gen	2.72e+02
13	42	63.6	193	25	W23619 Prolactin antagonist	3.40e+02
14	42	63.6	225	1	R92078 Recombinant rat prepr	3.40e+02
15	42	63.6	226	3	R14599 Rat prolactin	3.40e+02
16	42	63.6	392	2	R12345 Toxoplasma gondii pro	3.40e+02
17	42	63.6	428	2	R12352 Toxoplasma gondii P66	3.40e+02
18	42	63.6	473	22	W14005 Human SHC protein.	3.40e+02

19	42	63.6	473	15	R84637 Shc protein.	3.40e+02
20	42	63.6	474	17	R97243 SHC phosphotyrosine b	3.40e+02
21	42	63.6	494	15	R76615 Saccharomyces cerevis	3.40e+02
22	42	63.6	494	6	R31888 Defective tyrosine ki	3.40e+02
23	42	63.6	494	11	R58519 Protein kinase (HRR25	3.40e+02
24	41	62.1	570	26	W36002 Human Fcnd531 gene pr	4.26e+02
25	41	62.1	1719	16	R92100 Human RIZ allele D283	4.26e+02
26	41	62.1	3119	29	W44743 Mouse huntingtin prot	4.26e+02
27	41	62.1	3119	26	W36888 Mouse huntingtin's di	4.26e+02
28	40	60.6	101	18	R90022 Ub17 ubiquitin-lytic	5.32e+02
29	40	60.6	240	4	R22597 Foetal oncogene Pem s	5.32e+02
30	40	60.6	463	25	W19394 Human calpastatin (cl	5.32e+02
31	40	60.6	673	1	R05701 Human calpastatin pol	5.32e+02
32	40	60.6	706	18	R98925 Human APLP2.	5.32e+02
33	40	60.6	706	9	R47499 Human APLP2.	5.32e+02
34	40	60.6	763	10	R53778 Sequence of human amy	5.32e+02
35	40	60.6	1350	17	R91296 Drosophila nitric oxi	5.32e+02
36	40	60.6	1618	5	R27205 Human nestin.	5.32e+02
37	40	60.6	1618	11	R60127 Human nestin protein	5.32e+02
38	40	60.6	2205	15	R79048 Infectious rubella vi	5.32e+02
39	40	60.6	5035	5	R25450 MH mutant porcine rya	5.32e+02
40	39	59.1	740	5	R27530 Plasmodium falciparum	6.62e+02
41	39	59.1	858	1	P80806 Sequence of env prote	6.62e+02
42	39	59.1	891	1	P82677 ENVRN sequence from H	6.62e+02
43	39	59.1	957	11	R56990 Bacillus deramificans	6.62e+02
44	39	59.1	1717	27	W23331 Neuroblastoma indicat	6.62e+02
45	39	59.1	1831	27	W23329 Microtubule-associate	6.62e+02

ALIGNMENTS

RESULT 1
ID W44742 standard; Protein; 3144 AA.
AC W44742;

DT 01-JUN-1998 (first entry)

DE Human huntingtin gene; Huntington's disease; chromosome; marker;

KW Human; huntingtin gene; Huntington's disease; chromosome; marker;

OS Homo sapiens.

PD 02-DEC-1997.

PF 30-MAY-1995; 453265.

PR 20-MAY-1994; US-246982.

PR 05-MAR-1993; US-027498.

PR 01-JUL-1993; US-085000.

PR 30-MAY-1995; US-453265.

PA (GHEO) GEN HOSPITAL CORP.

PI Ambrose CM, Duyao MP, Gusella JF, MacDonald ME;

DR WPI: 98-031815/03.

DR N-PSDB; V05828.

PT Huntingtin protein and related nucleic acid - for diagnosis or

PS Claim 2; Fig 4; 112pp; English.

CC This is the amino acid sequence of the human huntingtin protein.

CC The gene sequence is characterised in that it contains a number of

CC CAG repeats in the 5' region (in this case 23 repeats). In healthy

CC individuals the usual number of CAG repeats in the gene is 11-34,

CC whereas in patients suffering from Huntington's disease (HD), the

CC number of repeats increases to 37-73 or 37-86. The huntingtin gene

CC spans 210 kb and encodes a protein of 348 kD. The gene is found in a

CC 500 kb region between the chromosomal markers D4S180 and D4S182 and is

CC preferentially mapped to the locus 4p16.3. The protein or the gene

CC encoding it, is useful for detecting a predisposition to develop HD,

CC for diagnosis and treatment of HD, especially by antisense and gene

CC therapy.

SQ Sequence 3144 AA;

Query Match 68.2%; Score 45; DB 29; Length 3144;

Best Local Similarity 60.0%; Pred. No. 1.72e+02;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2343 eeeevdpt 2352

|||||

Oy 1 EQQEVPPDT 10

RESULT 2

ID R58777 standard; Protein; 3144 AA.

AC R58777;

DE 13-APR-1995 (first entry)

DE Protein encoded by Huntingtin DNA; IT15 gene.

KW Polymerase chain reaction; primer; PCR; amplif; Huntingtin; IT15;

KW open reading frame; polymorphic; (CAG)n; trinucleotide repeat;

KW allele; Huntington's disease; HD; chromosome; juvenile HD; unstable;

KW expandable; linkage disequilibrium.

OS Homo sapiens.

PN EP-614977-A.

PD 14-SEP-1994. 301587.

PF 07-MAR-1994. 301587.

PR 05-MAR-1993; US-027498.

PR 01-JUL-1993; US-085000.

PA (GEHO) GEN HOSPITAL CORP.

PI Ambrose CM, Duyao MP, Gusella JF, MacDonald ME;

DR WPI: 94-281205/35.

DR N-PSDB: Q67401.

PT New huntingtin protein and related nucleic acid, antibodies etc.

PT - for treatment and diagnosis of neuro-degenerative disease,

PT specifically Huntington's disease

PS Claim 1; Page 22-33; 69pp; English.

CC This sequence is encoded by a large gene, termed "Huntingtin" or

CC "IT15" which is present in the proximal part of the 0.5 mb segment

CC between members D4S180 and D4S182 on Chromosome 4. The IT15 gene

CC spans about 210 kb and encodes a protein of approx. 348 kD. The

CC huntingtin open reading frame contains a polymorphic (CAG)n

CC trinucleotide repeat with at least 17 alleles in the normal

CC population, varying from about 11 to about 34 CAG copies. On

CC Huntington's disease (HD) chromosome, the length of the CAG repeat

CC is substantially increased, with between 37 to at least 73 copies.

CC This shows an apparent correlation with age of onset, the longest

CC segments are detected in juvenile HD cases. The presence of an

CC unstable, expandable trinucleotide repeat on HD chromosomes in the

CC region of the strongest linkage disequilibrium with the disorder

CC suggest that this alteration underlies the dominant phenotype of HD.

SQ Sequence 3144 AA;

Query Match 58.2%; Score 45; DB 11; Length 3144;

Best Local Similarity 60.0%; Pred. No. 1.72e+02;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2343 eeeevdnt 2352

Oy 1 EQQEVPPDT 10

RESULT 3

ID W36887 standard; Protein; 3144 AA.

AC W36887;

DE 13-MAR-1998 (first entry)

DE Previously undescribed protein encoded by a novel huntingtin (IT15) gene.

KW Huntingtin gene; IT15 gene; Huntington's disease; trinucleotide repeat;

KW neurodegenerative disorder; HD; gene therapy.

OS Homo sapiens.

PN US568288-A.

PD 11-NOV-1997. 246982.

PF 20-MAR-1994. 246982.

PR 05-MAR-1993; US-027498.

PR 01-JUL-1993; US-085000.

PA (GEHO) GEN HOSPITAL CORP.

PI Ambrose CM, Duyao MP, Gusella JF, MacDonald ME;

DR WPI: 97-358144/51.

DR N-PSDB: T97924.

PT Nucleic acid encoding huntingtin protein - useful for gene therapy

PT of Huntington's disease

PS Claim 2; Fig 4; 112pp; English.

CC The present sequence represents a previously undescribed protein,

CC encoded by a novel gene, termed huntingtin or IT15. The huntingtin

CC reading frame contains a polymorphic (CAG)n trinucleotide repeat with at

CC least 17 alleles in the normal population, varying from about 11 to 34

CC CAG copies. Huntington's disease (HD) is a progressive neurodegenerative

CC disorder characterised by motor disturbance, cognitive loss and

CC psychiatric manifestations. The genetic defect causing HD is assigned to

CC chromosome 4. On HD chromosomes, the length of the trinucleotide CAG

CC repeat is substantially increased, e.g. about 37 to at least 73 copies.

CC The huntingtin gene and proteins encoded by it, may be used for the

CC diagnosis or treatment of Huntington's disease. The huntingtin gene

CC is especially used in gene therapy of a symptomatic or presymptomatic

CC patient. The method comprises providing a functional huntingtin gene with

CC a (CAG)n repeat of the normal range of 11-34 copies, or an antisense

CC sequence, to the desired cells of the patient, in a manner that permits

CC the expression of the huntingtin protein provided by the gene, or

CC inhibits expression of the mutated huntingtin gene, for a time and in a

CC quantity sufficient to provide the huntingtin function to the cells of

CC the patient.

SQ Sequence 3144 AA;

Query Match 58.2%; Score 45; DB 26; Length 3144;

Best Local Similarity 60.0%; Pred. No. 1.72e+02;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2343 eeeevdnt 2352

Oy 1 EQQEVPPDT 10

RESULT 4

ID W09871 standard; Protein; 3144 AA.

AC W09871;

DE 27-JUL-1997 (first entry)

DE Human huntingtin.

KW Huntingtin associated protein-1; HAP1; Huntington's disease.

OS Homo sapiens.

EH Key Location/Qualifiers

FT Protein 1..230

FT W09717443-A1.

PN 15-MAY-1997.

PF 08-NOV-1996; U17858.

PR 09-NOV-1995; US-556419.

PA (UICJ) UNIV JOHNS HOPKINS.

PI Lanahan A, Li S, Li X, Ross CA, Sharp AH, Snyder S;

PI Worley PF;

DR WPI: 97-281032/25.

PT Determination of the binding of huntingtin to huntingtin-associated

PT protein-1 - useful for screening for drugs for treating or

PT preventing Huntington's disease

PS Claim 20; Page 22-32; 69pp; English.

CC Human huntingtin (Hn) polypeptide (W09871) is the product of the

CC Huntington's disease (HD) locus. Proteins which specifically bind

CC to Hn, such as human huntingtin associated protein-1 (HAP1) (see

CC also W09870), can be used in assays for screening drug candidates.

CC The binding between Hn and HAP1 is enhanced by an expanded

CC polyglutamine repeat in Hn, the length of which correlates with the

CC time of disease onset. HAP1, in contrast to Hn, is expressed

CC selectively in the brain, suggesting that it may contribute to the

CC brain-specific pathology of HD. Hn, or portions of it, esp. amino

CC acids 1-230, or yeast cells expressing Hn, can be used to identify

CC cpds. that bind to, displace or prevent binding of Hn and HAP1.

CC The method is useful for screening candidate drugs for treating,

CC delaying onset of, or preventing HD.

SQ Sequence 3144 AA;

Query Match 58.2%; Score 45; DB 22; Length 3144;

Best Local Similarity 60.0%; Pred. No. 1.72e+02;

Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2343 eeeevdnt 2352

Oy 1 EQQEVPPDT 10

```

RESULT 5
ID R05669 standard; peptide; 140 AA.
AC R05669;
DT 15-AUG-1990 (first entry)
DE Gamma-chicken atrial natriuretic peptide.
KW Gamma-chicken atrial natriuretic peptide; diuretic; hypertensive.
OS Gallus sp.
FH Key Location/Qualifiers
FT disulfide_bond 118..134
FT region 25..140
FT /label=Gamma-chANP specific.
PN J02025499-A.
PD 26-JAN-1990.
PF 14-JUL-1988; 173739.
PR 14-JUL-1988; JP-173739.
PA (MATSU) Matsuo T.
DR WPI: 90-071804/10.
DR N-PSDB; Q03465.
PT Physio-active peptide derived from birds - has cysteine bridge,
PT specified amino acid sequence and diuretic and hypertensive properties.
PS Disclosure; Fig 13; 16pp; Japanese.
CC Gamma-chANP is obtained from the 12 kD fraction obtained from treated
CC homogenised chicken heart tissue.
CC See also Q03466-Q03468, R03301 and R03302.
SQ Sequence 140 AA;

Query Match 65.2%; Score 43; DB 1; Length 140;
Best Local Similarity 44.4%; Pred. No. 2.72e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 66 qtgeevppe 74
QY 1 EQQEVPPD 9

RESULT 6
ID R83055 standard; Protein; 414 AA.
AC R83055;
DT 25-JUN-1996 (first entry)
DE Transforming growth factor-beta 2.
KW macrophage inducible nitric oxide synthase; iNOS; constitutive NOS;
KW interleukin-1-beta; transforming growth factor-beta; TGF-beta; IL1-beta;
KW nitric oxide production; hypotension; inflammation; septic shock;
KW treatment.
OS Mammalian sp.
FH Key Location/Qualifiers
FT protein 303..414
FT /note= "represents the mature active TGF beta-1 mol."
PN W09526745-A1.
PD 12-OCT-1995.
PF 05-APR-1994; U03705.
PR 05-APR-1994; WO-U03705.
PA (HARD ) HARVARD COLLEGE.
PI Lee M, Perrella MA;
DR WPI: 95-35843/46.
DR N-PSDB; T05877.
PT Treatment of hypotension, esp. in septic shock - by administering
PT transforming growth factor-beta e.g. to inhibit inducible nitric
PT oxide synthase gene transcription
PS Disclosure; Fig 18; 52pp; English.
CC Transforming growth factor-beta 2 (TGF-beta 2) has been found to inhibit
CC inducible nitric oxide synthase (iNOS) gene transcription, esp. in
CC interleukin-1-beta (IL1-beta) stimulated rat smooth muscle cells, and at
CC a dose which does not inhibit constitutive NOS. TGF-beta 1 (R83054) or 2
CC or their active fragments (esp. derived from the carboxy-terminal 112
CC amino acids), can be used in the treatment of hypotension, such as that
CC associated with severe inflammation or septic shock.
SQ Sequence 414 AA;

Query Match 65.2%; Score 43; DB 16; Length 414;
Best Local Similarity 62.5%; Pred. No. 2.72e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 59 epeevppe 66
QY 2 EQQEVPPD 9

RESULT 7
ID R73597 standard; Protein; 414 AA.
AC R73597;
DT 20-DEC-1995 (first entry)
DE Human TGF-beta 2 protein.
KW Transforming growth factor-beta; Human TGF-beta protein; TGF-beta 1;
KW TGF-beta 2; TGF-beta 3; osteogenic cell source; OCS; bone deficiency;
KW bone-inducing cofactor.
OS Homo sapiens.
PN US5409896-A.
PD 25-APR-1995.
PF 12-NOV-1993; 401906.
PR 01-SEP-1989; US-401906.
PR 12-NOV-1991; US-790856.
PR 18-MAY-1993; US-063841.
PR 12-NOV-1993; US-132405.
PA (GETH ) GENENTECH INC.
PI Ammann AJ, Rudman CG;
DR WPI: 95-169610/22.
PT Compsn. for treating skeletal tissue deficiency - comprising
PT transforming growth factor-beta and an osteogenic cell source in a
PT carrier
PS Disclosure; Column 17-20; 19pp; English.
CC This sequence represents human transforming growth factor-beta 2
CC (TGF-beta 2). The sequences for human TGF-beta 1 (see R73596) and human
CC TGF-beta 3 (see R73598) are claimed within the scope of the invention.
CC The invention is a composition consisting of a TGF-beta protein and an
CC osteogenic cell source (OCS) formulated in an acceptable carrier other
CC than a bone morphogenic cofactor. This composition can be used for the
CC restoration of bone deficiency. This provides for the generation of a
CC mature bone only where it is required, without the inclusion of a
CC specific bone-inducing cofactor. This method can be used with any of the
CC 5 human TGF-beta's or with TGF-beta from other species.
SQ Sequence 414 AA;

Query Match 65.2%; Score 43; DB 14; Length 414;
Best Local Similarity 62.5%; Pred. No. 2.72e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 59 epeevppe 66.
QY 2 EQQEVPPD 9

RESULT 8
ID P91899 standard; Protein; 442 AA.
AC P91899;
DT 23-DEC-1990 (first entry)
DE Sequence encoded by human transforming growth factor (TGF) beta-2
DE precursor 442 cDNA in pPC-21
KW Cell differentiation; cell proliferation.
OS Homo sapiens.
FH Key Location/Qualifiers
FT region 20..442
FT /note="Claimed"
FT peptide 4..19
FT /note="Signal"
FT cleavage_site 20..21
FT region 116..144
FT /note="This entire SQ is replaced with Asn in simian
FT TGF-beta-2-414"
FT modified_site 72
FT /note="Potential glycosylation site"
FT modified_site 168
FT /note="Potential glycosylation site"
FT modified_site 269
FT /note="Potential glycosylation site"

```

FT protein 331..442
 PN DE3833897-A.
 PD 03-MAY-1989.
 PF 05-OCT-1988; 833897.
 PR 18-AUG-1988; US-234065.
 PA (ONCO-) Oncogen.
 PI Purchio AF, Madisen L, Webb N;
 DR WPI; 89-138796/19.
 DR N-PSDB; N90767.
 PT New DNA sequence encoding transforming growth factor beta 2 -
 used for large scale expression in eucaryotic cells
 PS Claim 4; Fig 1a; 2pp; German.
 CC PolyA-RNA was isolated from the tamoxifen-treated, human prostatic
 adenocarcinoma line PC-3 and converted to cDNA. TGF DNA is pref. used
 for control of the SV40 promoter. and expressed in CHO cells. The simian
 CC SQ is also claimed.
 SQ Sequence 442 AA;
 Query Match 65.2%; Score 43; DB 1; Length 442;
 Best Local Similarity 62.5%; Pred. No. 2.72e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 59 epeevppe 66
 | :||||:
 Qy 2 EQQEVPPD 9

RESULT 9
 ID R79922 standard; Protein; 442 AA.
 AC R79922;
 DT 28-MAY-1996 (first entry)
 DE Human transforming growth factor-2.
 KW TGF-beta1; TGF-beta2; transforming growth factor; protein;
 KW cell differentiation; cell proliferation; CHO; Chinese hamster;
 KW ovary; COS; monkey kidney; animal; mammal.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT peptide 4..19
 FT cleavage_site 20..21
 FT /note= "signal peptide"
 FT /note= "putative signal sequence cleavage site"
 FT peptide 331..442
 FT /note= "mature peptide"
 PN EP-676474-A1.
 PD 11-OCT-1995.
 PF 14-DEC-1989; 104223.
 PR 16-DEC-1988; US-285140.
 PR 05-DEC-1989; US-446020.
 PA (ONCO) ONCOGEN LP.
 PI Madisen L, Purchio AF, Webb N;
 DR WPI; 95-346094/45.
 DR N-PSDB; T04116.
 PT Hybrid transforming growth factor beta-1/TGF-beta-2 precursor - used
 to produce biologically active, mature TGF-beta-2
 PS Disclosure; Fig.1a; 52pp; English.
 CC This sequence is expressed in a host cell, preferably a
 CC COS or CHO cell, so the host cell produces active TGF-beta2. The
 CC produced TGF-beta2 protein can be used to regulate cellular
 CC differentiation and proliferation.
 SQ Sequence 442 AA;
 Query Match 65.2%; Score 43; DB 16; Length 442;
 Best Local Similarity 62.5%; Pred. No. 2.72e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 59 epeevppe 66
 | :||||:
 Qy 2 EQQEVPPD 9

RESULT 10
 ID R20125 standard; Protein; 442 AA.
 AC R20125;

FT protein 331..442
 PN DE3833897-A.
 PD 03-MAY-1989.
 PF 05-OCT-1988; 833897.
 PR 18-AUG-1988; US-234065.
 PA (ONCO-) Oncogen.
 PI Purchio AF, Madisen L, Webb N;
 DR WPI; 89-138796/19.
 DR N-PSDB; N90767.
 PT New DNA sequence encoding transforming growth factor beta 2 -
 used for large scale expression in eucaryotic cells
 PS Claim 4; Fig 1a; 2pp; German.
 CC PolyA-RNA was isolated from the tamoxifen-treated, human prostatic
 adenocarcinoma line PC-3 and converted to cDNA. TGF DNA is pref. used
 for control of the SV40 promoter. and expressed in CHO cells. The simian
 CC SQ is also claimed.
 SQ Sequence 442 AA;
 Query Match 65.2%; Score 43; DB 1; Length 442;
 Best Local Similarity 62.5%; Pred. No. 2.72e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 59 epeevppe 66
 | :||||:
 Qy 2 EQQEVPPD 9

RESULT 11
 ID R05748 standard; protein; 442 AA.
 AC R05748;
 DT 02-NOV-1990 (first entry)
 DE Human TGF-Beta2-442 precursor.
 KW Human TGF-Beta2 precursor; cancer; tumorcide; ss.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT protein 331..442
 FT peptide 4..19
 FT /label-Sinal peptide.
 PN EP-376785-A.
 PD 4-JUL-1990.
 PF 14-DEC-1989; 403480.
 PR 16-DEC-1988; US-285140.
 PR 5-DEC-1989; US-446020.
 PA (ONCO-) Oncogen Ltd Partner.
 PI Purchio AF, Madisen L, Webb N;
 DR WPI; 90-203127/27.
 DR N-PSDB; Q05126
 PT Cloning and expression of transforming growth factor beta 2 -
 used for treatment of tumors or for augmenting wound healing.
 PS Claim 1; Fig 1a; 58pp; English.
 CC TGF-Beta2 may be used in treatment of tumors at effective doses,
 CC and may also be useful in augmenting wound healing by stimulating
 CC cell proliferation. The growth factor can be produced at high
 CC levels from a CHO expression system.
 SQ Sequence 442 AA;
 Query Match 65.2%; Score 43; DB 1; Length 442;
 Best Local Similarity 62.5%; Pred. No. 2.72e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 59 epeevppe 66
 | :||||:
 Qy 2 EQQEVPPD 9

DT 16-APR-1992 (first entry)
 DE Sequence of human transforming growth factor (TGF) beta-2-442.
 KW Hypertension therapy; hypotensive agent; blood pressure modulator.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT peptide 4..19
 FT protein 330..442
 PN WO9119513-A.
 PD 26-DEC-1991.
 PF 20-JUN-1991; U04449.
 PR 20-JUN-1990; US-541221.
 PA (BRIM) BRISTOL-MYERS SQUIB.
 PI Oleson FB, Comerreski CR;
 DR WPI; 92-024199/03.
 DR N-PSDB; Q20290.
 PT Use of transforming growth factor (TGF)-beta and their
 PT antagonists - for modulating blood pressure, for treating
 PT hypertension and hypotension
 PS Disclosure; Fig 2; 42pp; English.
 CC A new method for treating hypertension comprises administering a
 CC transforming growth factor (TGF)-beta to an individual at a dose
 CC effective for lowering blood pressure; the TGF-beta may be e.g.
 CC mature TGF-beta, TGF-beta2, a mature TGF-beta2 hybrid, TGF-
 CC beta1 precursor, a latent TGF-beta2 precursor, hybrid TGF-beta1/TGF-
 CC beta2 precursor, a latent TGF-beta1 complex or a latent TGF-beta2
 CC complex.
 SQ Sequence 442 AA;
 Query Match 65.2%; Score 43; DB 4; Length 442;
 Best Local Similarity 62.5%; Pred. No. 2.72e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 59 epeevppe 66
 | :||||:
 Qy 2 EQQEVPPD 9

```

RESULT 12
ID R03663 standard; protein; 700 AA.
AC R03663;
DT 21-AUG-1990 (first entry)
DE Human myb related gene product.
KW Cancer; myb; myeloblast; ds.
OS Homo sapiens.
PN J02053486-A.
PD 22-FEB-1990.
PR 19-AUG-1988; 205907.
PR 19-AUG-1988; JP-205907.
PA (RIKA) Rikagaku Kenkyusho.
DR WPI; 90-103118/14.
DR N-PSDB; Q03741.
PT Human Myeloblast related gene -
PT coded by bases sequence including 105 adenine 1150 thymine, etc.
PS Claim 2; Fig 2; 8pp; Japanese.
CC Proteins generated by the sequence may be used to raise antibodies useful
CC in determining the copy number of the cancer gene.
SQ Sequence 700 AA;

Query Match 65.2%; Score 43; DB 1; Length 700;
Best Local Similarity 40.0%; Pred. No. 2.72e+02;
Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 277 ddqegsppet 286
::: ||:|
QY 1 EEQVEVPPDT 10

RESULT 13
ID W23619 standard; protein; 193 AA.
AC W23619;
DT 11-FEB-1998 (first entry)
DE Prolactin antagonist (substituted rat prolactin).
KW Prolactin antagonist; phosphorylation; hyperprolactinaemia;
KW prolactinoma; prostate cancer; tumour; T-lymphoma; infertility;
KW lactation; miscarriage; ovulation; antibody; therapy; rat.
OS Rattus sp.
OS Synthetic.
FH Key Location/Qualifiers
FT Misc_difference 173 /note= "variable site"
FT FT
PN W09727865-A1.
PD 07-AUG-1997.
PF 30-JAN-1997; U01435.
PR 31-JAN-1996; US-594809.
PA (REGC ) UNIV CALIFORNIA.
PI Walker AM.
DR WPI; 97-402308/37.
PT Substituted prolactin peptide(s) and proteins having an amino acid
PT substitution for serine in the C-terminal - useful as prolactin
PT antagonists, e.g. for treating prolactin dependent cancers
PS Claim 4; Page 101-102; 158pp; English.
CC This protein comprises rat prolactin, substituted at residue 173
CC (serine in the native sequence). It has prolactin antagonist
CC activity, antagonising the stimulation of T lymphoma cell growth in
CC the presence of non-phosphorylated prolactin. Other claimed
CC prolactin antagonists (see W23608-18) comprise prolactin
CC substitution mutant C-terminal peptides. Claimed antagonists can
CC be used for the treatment of prolactin dependent cancers and can
CC inhibit T-lymphoma cell proliferation. They are also useful for
CC treatment of prolactinoma, infertility related to abnormal
CC and ovulation irregularities, as well as in assays to measure
CC levels of non-phosphorylated and phosphorylated prolactin as an
CC indicator of reproductive pathologies and presence or status of a
CC prolactin-dependent tumour, and to raise polyclonal and monoclonal
CC antibodies.
SQ Sequence 193 AA;

```

```

Query Match 63.6%; Score 42; DB 25; Length 193;
Best Local Similarity 55.6%; Pred. No. 3.40e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 64 eqagkvpe 72
::: |||:
QY 1 EEQVEVPPD 9

```

```

RESULT 14
ID P82078 standard; protein; 225 AA.
AC P82078;
DT 18-OCT-1990 (first entry)
DE Recombinant rat preprolactin.
KW Prolactin; milk; contraceptive; dairy cows; lactation.
OS Rattus rattus.
FH Key Location/Qualifiers
FT protein 1..225 /label=preprolactin
FT protein 29..225 /label=prolactin
FT FT
PN US4725549-A.
PD 16-FEB-1988.
PF 23-MAR-1984; 592714.
PR 22-SEP-1980; US-189160.
PR 23-MAR-1984; US-592714.
PA (REGC) University of California.
PI Cooke NE; Baxter JD;
DR WPI; 88-070922/10.
DR N-PSDB; N80114.
PT DNA coding for prolactin - obt'd. by prep'n. of reverse transcript
PT of mRNA coding for prolactin and inserting into a transfer vector.
PS Disclosure; p; English.
CC The cDNA encoding the prolactin can be inserted into expression vectors
CC for the prodn. of rat prolactin which can be admin. to dairy cows to
CC increase milk yield. The protein can also be used as a female
CC contraceptive and to ensure adequate milk prodn. for breast feeding
CC mothers.
CC See also P82079.
SQ Sequence 225 AA;

```

```

Query Match 63.6%; Score 42; DB 1; Length 225;
Best Local Similarity 55.6%; Pred. No. 3.40e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 96 eqagkvpe 104
::: |||:
QY 1 EEQVEVPPD 9

```

```

RESULT 15
ID R14599 standard; protein; 226 AA.
AC R14599;
DT 21-JAN-1992 (first entry)
DE Rat prolactin.
KW Recombinant DNA.
OS Rattus rattus.
PN J03219876-A.
PD 27-SEP-1991.
PF 24-JAN-1990; 014511.
PR 24-JAN-1990; JP-014511.
PA (SHIK-) SHIKISHIMA BOSEKI K.
DR WPI; 91-329111/45.
DR N-PSDB; Q14451, Q14452.
PT Mass-prodn. of rat prolactin-producing recombinant DNA - is by
PT integration of promoter, Shine-Dalgarno sequence and translation
PT initiation codon upstream of table 1 coding gene.
PS Disclosure; Table 1; 12pp; Japanese.
CC The gene encoding the protein can be ligated into an expression
CC plasmid with a promoter, SD sequence and initiation codon for the
CC prodn. of recombinant rat prolactin. The protein is useful for the
CC study of the physiological activity of rat and human prolactin.
SQ Sequence 226 AA;

```

Query Match 63.6%; Score 42; DB 3; Length 226;
Best Local Similarity 55.6%; Pred. No. 3.40e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 97 eqagkvppe 105
1: | | | | |
QY 1 EEQOEVPDP 9

Search completed: Thu May 13 15:31:55 1999
Job time : 17 secs.

W P S R L H
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu May 13 15:32:14 1999; MasPar time 2.87 Seconds
Tabular output not generated. 130.735 Million cell updates/sec

Title: >US-09-040-485-7
Description: (1-10) from US09040485.pap
Perfect Score: 66
Sequence: 1 EEQVEVPDPT 10

Scoring table: PAM 150
Gap 15

Searched: 116738 seqs, 37463448 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pirs8
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 21.325; Variance 28.553; scale 0.747

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match %	Description	Pred. No.
1	65	98.5	peptide-aspartate bet	1.90e-03
2	65	98.5	aspartyl beta-hydroxy	1.90e-03
3	55	83.3	probable membrane pro	2.96e-01
4	49	74.2	prolactin - golden ha	4.98e+00
5	47	71.2	single stranded DNA-b	1.22e+01
6	46	69.7	heat shock protein x4	1.90e+01
7	46	69.7	hypothetical protein	1.90e+01
8	46	69.7	elastic titin - human	1.90e+01
9	45	68.2	neural UNC-119 prot	2.93e+01
10	45	68.2	flagellar antigen - T	2.93e+01
11	45	68.2	histidine-rich calciu	2.93e+01
12	45	68.2	Huntington disease-as	2.93e+01
13	45	68.2	ankyrin 2, neuronal 1	2.93e+01
14	44	66.7	hypothetical protein	4.48e+01
15	44	66.7	uroporphyrinogen deca	4.48e+01
16	44	66.7	hypothetical protein	4.48e+01
17	44	66.7	nucleolar protein NOP	4.48e+01
18	44	66.7	transforming protein	4.48e+01
19	43	65.2	alpha-atrial natriure	6.83e+01
20	43	65.2	hypothetical protein	6.83e+01
21	43	65.2	coat protein - phase	6.83e+01
22	43	65.2	arrestin - bovine	6.83e+01
23	43	65.2	cyclin A - chicken	6.83e+01

ALIGNMENTS

RESULT 1		BABOH		#type complete	
ENTRY		peptide-aspartate beta-dioxygenase (EC 1.14.11.16) - bovine		aspartyl (asparaginyl) beta-hydroxylase	
TITLE		aspartyl (asparaginyl) beta-hydroxylase			
ALTERNATE_NAMES		#formal_name Bos primigenius taurus #common_name cattle			
ORGANISM		31-Dec-1993 #sequence_revision 10-Feb-1995 #text_change			
DATE		29-May-1998			
ACCESSIONS		A42969; A39470; B39470; C39470; S27948			
REFERENCE		A42969			
#authors		Jia, S.; Vandusen, W.J.; Diehl, R.E.; Kohl, N.E.; Dixon, R.A.; Elliston, K.O.; Stern, A.M.; Friedman, P.A.			
#journal		J. Biol. Chem. (1992) 267:14322-14327			
#title		cDNA cloning and expression of bovine aspartyl (asparaginyl) beta-hydroxylase.			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			
#cross-references		EMBL:M91213; NID:g162693; PID:g162694			
#molecule_type		beta-hydroxylase			
#accession		A42969			
#residues		1-754			

```

KEYWORDS
FEATURE
2-56      #domain intracellular #status predicted #label INC\
57-78      #domain transmembrane #status predicted #label TRM\
289-754    #product peptide-aspartate beta-dioxygenase, 56K form
311-754    #status predicted #label 56K\
          #product peptide-aspartate beta-dioxygenase, 52K form
337-370    #domain tetratricopeptide repeat homology #label TTR\
371-404    #domain tetratricopeptide repeat homology #label TR2\
13,96,466,702 #binding_site carbohydrate (Asn) (covalent) #status
              predicted
SUMMARY    #length 754 #molecular-weight 84998 #checksum 9667

Query Match      98.5%; Score 65; DB 1; Length 754;
Best Local Similarity 90.0%; Pred. No. 1.90e-03;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 300 EEOQEVPPET 309
      |||||
QY 1 EEOQEVPPDT 10

RESULT 2
ENTRY
TITLE      #type complete
ORGANISM    aspartyl beta-hydroxylase - human
DATE        #formal_name Homo sapiens #common_name man
29-May-1998 #sequence_revision 29-May-1998 #text_change
10-Jul-1998
ACCESSIONS  I38423
REFERENCE     I38423
#authors      Koriath, F.; Gieffers, C.; Frey, J.
#journal      Gene (1994) 150:395-399
#title        Cloning and characterization of the human gene encoding
              aspartyl beta-hydroxylase.
#cross-references MUID:95121937
#accession    I38423
#status       preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues     1-757 #label RES
#cross-references EMBL:U03109; NID:G458031; PID:G458032
CLASSIFICATION #superfamily peptide-aspartate beta-dioxygenase;
              tetratricopeptide repeat homology
FEATURE
54-75      #domain transmembrane #status predicted #label TRM
SUMMARY    #length 757 #molecular-weight 85498 #checksum 2143

Query Match      98.5%; Score 65; DB 2; Length 757;
Best Local Similarity 90.0%; Pred. No. 1.90e-03;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 304 EEOQEVPPET 313
      |||||
QY 1 EEOQEVPPDT 10

RESULT 3
ENTRY
TITLE      #type complete
ALTERNATE_NAMES probable membrane protein YMR160w - yeast (Saccharomyces
                  cerevisiae)
ORGANISM      hypothetical protein YMR8520.09
DATE          #formal_name Saccharomyces cerevisiae
08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change
12-Dec-1997
ACCESSIONS    S54518; S54605
REFERENCE     S54518
#authors      Hunt, S.; Bowman, S.
#submision    submitted to the EMBL Data Library, May 1995
#accession    S54518
#molecule_type DNA
#residues     1-816 #label HUN
#cross-references GB:249705; EMBL:249700; NID:G825556; PID:G825565;
              EMBL:249705; MIPS:YMR160w

```

```

GENETICS    ##experimental_source strain AB972
FEATURE      #map_position 13R
KEYWORDS     transmembrane protein
328-344      #domain transmembrane #status predicted #label TMM
SUMMARY      #length 816 #molecular-weight 95096 #checksum 3162

Query Match      83.3%; Score 55; DB 2; Length 816;
Best Local Similarity 77.8%; Pred. No. 2.96e-01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 24 EDOQEVPPQ 32
      |||||
QY 1 EEOQEVPPD 9

RESULT 4
ENTRY
TITLE      #type complete
ORGANISM    prolactin - golden hamster
DATE        #formal_name Mesocricetus auratus #common_name golden hamster
21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change
16-Feb-1997
ACCESSIONS    A49159
REFERENCE     A49159
#authors      Southard, J.N.; Sanchez-Jimenez, F.; Campbell, G.T.;
              Talamantes, F.
#journal      Endocrinology (1991) 129:2965-2971
#title        Sequence and expression of hamster prolactin and growth
              hormone messenger RNAs.
#cross-references MUID:92063850
#accession    A49159
#status       preliminary
#molecule_type mRNA
#residues     1-226 #label SOU
#note         sequence extracted from NCBI backbone (NCBIN:66296,
              NCBIP:66298)
CLASSIFICATION #superfamily prolactin
FEATURE
33-38,85-201,
218-226      #disulfide_bonds #status predicted
SUMMARY      #length 226 #molecular-weight 25582 #checksum 3394

Query Match      74.2%; Score 49; DB 2; Length 226;
Best Local Similarity 66.7%; Pred. No. 4.98e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 97 EEOQVPPPE 105
      |||||
QY 1 EEOQEVPPD 9

RESULT 5
ENTRY
TITLE      #type complete
ORGANISM    single stranded DNA-binding protein - Aquifex aeolicus
DATE        #formal_name Aquifex aeolicus
21-Aug-1998
ACCESSIONS    H70305
REFERENCE     H70305
#authors      Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
              Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;
              Keller, M.; AuJay, M.; Huber, R.; Feldman, R.A.; Short,
              J.M.; Olson, G.J.; Swanson, R.V.
#journal      Nature (1998) 392:353-358
#title        The complete genome of the hyperthermophilic bacterium
              Aquifex aeolicus.
#cross-references MUID:98196666
#accession    H70305
#status       preliminary; nucleic acid sequence not shown;
              translation not shown
#molecule_type DNA
#residues     1-147 #label AQF

```

##cross-references GB:AE000672; NID:g2982810; PID:g2982816; GB:AE000657
 ##experimental_source strain VF5

GENETICS
 #gene ssb
 CLASSIFICATION #superfamily bacterial single-stranded DNA-binding protein;
 single-stranded DNA-binding protein homology
 SUMMARY #length 147 #molecular-weight 17132 #checksum 2119

Query Match 71.2%; Score 47; DB 2; Length 147;
 Best Local Similarity 75.0%; Pred. No. 1.22e+01;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 115 EEEVPP 122
 ||:||||
 QY 1 EEQVEVPP 8

RESULT 6
 ENTRY C22175 #type fragment
 TITLE heat shock protein X4 - African clawed frog (fragment)
 ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
 DATE 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change
 24-Sep-1998

ACCESSIONS C22175
 REFERENCE A22175
 #authors Bienz, M.
 #journal Proc. Natl. Acad. Sci. U.S.A. (1984) 81:3138-3142
 #title Developmental control of the heat shock response in Xenopus.
 #cross-references MIM:84221917
 #accession C22175
 #status preliminary
 #molecule_type mRNA
 #residues 1-68 #label BIE

CLASSIFICATION #superfamily alpha-crystallin
 SUMMARY #length 68 #checksum 9585
 Query Match 69.7%; Score 46; DB 2; Length 68;
 Best Local Similarity 55.6%; Pred. No. 1.90e+01;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 48 DAQEIPPD 56
 : ||:||||
 QY 2 EEQVEVPPDT 10

RESULT 7
 ENTRY S50808 #type complete
 TITLE hypothetical protein YJL065c - yeast (Saccharomyces
 Cerevisiae)
 ALTERNATE_NAMES hypothetical protein HRD167; hypothetical protein J1115
 ORGANISM #formal_name Saccharomyces cerevisiae
 DATE 13-Jan-1995 #sequence_revision 08-Sep-1995 #text_change
 21-Nov-1997
 ACCESSIONS S50808: S47127; S56839
 REFERENCE S50798
 #authors Vandenbol, M.; Durand, P.; Dion, C.; Portetelle, D.; Hilger,
 F.
 #journal Yeast (1995) 11:57-60
 #title Sequence of a 17.1 kb DNA fragment from chromosome X of
 Saccharomyces cerevisiae includes the mitochondrial
 ribosomal protein L8.

#accession S50808
 #status nucleic acid sequence not shown; translation not shown
 #molecule_type DNA
 #residues 1-167 #label VAN
 ##cross-references EMBL:234288; NID:g498992; PID:g499003
 ##note The nucleotide sequence was submitted to the EMBL Data
 Library, June 1994

REFERENCE S47117
 #authors Vandenbol, M.; Durand, P.; Dion, C.; Portetelle, D.; Hilger,
 F.
 #submission submitted to the EMBL Data Library, June 1994

#description Sequence analysis of a 17.1 kb DNA fragment from chromosome X
 of Saccharomyces cerevisiae includes the mitochondrial
 ribosomal protein L8.

#accession S47127
 #molecule_type DNA
 #residues 1-167 #label VAW
 ##cross-references EMBL:234288; NID:g498992; PID:g499003

REFERENCE S56835
 #authors Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.
 #submission submitted to the Protein Sequence Database, September 1995
 #accession S56839
 #molecule_type DNA
 #residues 1-167 #label POH

GENETICS
 ##cross-references EMBL:249340; NID:g1008212; PID:g1008213; MIPS:YJL065c

#map_position 10L
 SUMMARY #length 167 #molecular-weight 18792 #checksum 9768

Query Match 69.7%; Score 46; DB 2; Length 167;
 Best Local Similarity 62.5%; Pred. No. 1.90e+01;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 68 QQQQVPP 75
 : ||:||||
 QY 1 EEQVEVPP 8

RESULT 8
 ENTRY I38346 #type fragment
 TITLE elastic titin - human (fragment)
 ORGANISM #formal_name Homo sapiens #common_name man
 DATE 29-May-1998 #sequence_revision 29-May-1998 #text_change
 05-Jun-1998

ACCESSIONS I38346
 REFERENCE A57430
 #authors Labelit, S.; Kolmerer, B.
 #journal Science (1995) 270:293-296
 #title Titins: giant proteins in charge of muscle ultrastructure and
 elasticity.

#cross-references MIM:96026330
 #accession I38346
 #status preliminary; translated from GB/EMBL/DDBJ
 #molecule_type mRNA
 #residues 1-7962 #label RES
 ##cross-references EMBL:X90569; NID:g1017426; PID:g1017427

GENETICS
 #gene GDB:TTN
 ##cross-references GDB:127867; OMIM:188840
 #map_position 2q31-2q31
 SUMMARY #length 7962 #checksum 120

Query Match 69.7%; Score 46; DB 2; Length 7962;
 Best Local Similarity 62.5%; Pred. No. 1.90e+01;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 6815 EEEVPP 6822
 ||:||||
 QY 2 EEQVEVPPD 9

RESULT 9
 ENTRY JCS728 #type complete
 TITLE neuronal UNC-119 protein - Caenorhabditis briggsae
 ORGANISM #formal_name Caenorhabditis briggsae
 DATE 09-Dec-1997 #sequence_revision 09-Dec-1997 #text_change
 26-Feb-1998

ACCESSIONS JCS728
 REFERENCE JCS728
 #authors Maguro, M.; Pilgrim, D.
 #journal Gene (1996) 183:77-85
 #title Conservation of function and expression of unc-119 from two
 Caenorhabditis species despite divergence of non-coding
 DNA.

#accession JC5728
 ##molecule_type DNA
 ##residues 1-217 ##label MAD
 ##cross-references GB:U45326; NID:g1181702; PID:g1181703
 COMMENT This protein is involved in nervous system function.

GENETICS
 #gene unc-119
 #introns 13/3: 56/2; 173/1
 SUMMARY #length 217 #molecular-weight 25094 #checksum 7052

Query Match 68.2%; Score 45; DB 2; Length 217;
 Best Local Similarity 50.0%; Pred. No. 2.93e+01;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 4 EEOQSIIPGS 13
 ||:||||:
 QY 1 EEOQEVPPDT 10

RESULT 10
 ENTRY S47436 #type fragment
 TITLE flagellar antigen - Trypanosoma brucei (fragment)
 ORGANISM #formal_name Trypanosoma brucei
 DATE 13-Jan-1995 #sequence_revision 30-Jan-1998 #text_change
 30-Jan-1998
 ACCESSIONS S47436
 REFERENCE S47436
 #authors Imboden, M.; Mueller, N.; Hemphill, A.; Mattioli, X.Y.Z.; Seebeck, T.
 #submission submitted to the EMBL Data Library, August 1994
 #description Repetitive proteins from the flagellar cytoskeleton of African Trypanosomes are diagnostically useful antigens.

#accession S47436
 ##molecule_type mRNA
 ##residues 1-411 ##label IMB
 ##cross-references EMBL:236281; NID:g530358; PID:g530359
 ##experimental_source strain stock TREU 1285
 SUMMARY #length 411 #checksum 428

Query Match 68.2%; Score 45; DB 2; Length 411;
 Best Local Similarity 50.0%; Pred. No. 2.93e+01;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 10 EEOQVPAAE 19
 ||:||||:
 QY 1 EEOQEVPPDT 10

RESULT 11
 ENTRY A34373 #type complete
 TITLE histidine-rich calcium-binding protein precursor - rabbit
 ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic rabbit
 DATE 08-Jun-1990 #sequence_revision 08-Jun-1990 #text_change
 10-Sep-1997

ACCESSIONS A34373
 REFERENCE A34373
 #authors Hofmann, S.L.; Goldstein, J.L.; Orth, K.; Moomaw, C.R.; Slaughter, C.A.; Brown, M.S.
 #journal J. Biol. Chem. (1989) 264:18083-18090
 #title Molecular cloning of a histidine-rich Ca(2+)-binding protein of sarcoplasmic reticulum that contains highly conserved repeated elements.
 #cross-references MUID:90036884

#accession A34373
 ##status preliminary
 ##molecule_type mRNA
 ##residues 1-852 ##label HOF
 ##cross-references GB:J05080; NID:g165099; PID:g165100
 KEYWORDS calcium binding
 SUMMARY #length 852 #molecular-weight 96116 #checksum 3434

Query Match 68.2%; Score 45; DB 2; Length 852;

Best Local Similarity 55.6%; Pred. No. 2.93e+01;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 190 EEEEVSPSE 198
 ||:||||:
 QY 1 EEOQEVPPD 9

RESULT 12
 ENTRY A46068 #type complete
 TITLE Huntingdon disease-associated protein - human
 ORGANISM #formal_name Homo sapiens #common_name man
 DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
 29-Aug-1997

ACCESSIONS A46068
 REFERENCE A46068
 #authors MacDonald, M.E.; Ambrose, C.M.; Duyao, M.P.; Myers, R.H.; Lin, C.; Srinidhi, L.; Barnes, G.; Taylor, S.A.; James, M.; Groot, N.; MacFarlane, H.; Jenkins, B.; Anderson, M.A.; Wexler, N.S.; Gusella, J.F.; Bates, G.P.; Baxendale, S.; Hummerich, H.; Kirby, S.; North, M.; Youngman, S.; Mott, R.; Zehetner, G.; Sedlacek, Z.; Poustka, A.; Frischauf, A.M.; Buckler, A.J.; Church, D.; Doucette-Stamm, L.; O'Donovan, M.C.; Riba-Ramirez, L.; Shah, M.; Stanton, V.P.; Strobel, S.A.; Draths, K.M.; Wales, J.L.; Dervan, P.; Housman, D.E.; Altherr, M.; Shiang, R.; Thompson, L.; Fielder, T.; Wasmuth, J.J.; Tagle, D.; Valdes, J.; Elmer, L.; Allard, M.; Castilla, L.; Swaroop, M.; Blanchard, K.; Collins, F.S.; Snell, R.; Holloway, T.; Gillespie, K.; Datson, N.; Shaw, D.; Harper, P.S.

#journal Cell (1993) 72:971-983
 #title A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes.
 #accession A46068
 ##status preliminary
 ##molecule_type mRNA
 ##residues 1-3144 ##label MAC
 ##cross-references GB:L12392
 REFERENCE I54337
 #authors Lin, B.; Rommens, J.M.; Graham, R.K.; Kalchman, M.; Hayden, M.R.; MacDonald, H.; Nasir, J.; Delaney, A.; Goldberg, Y.P.; Hayden, M.R.
 #journal Hum. Mol. Genet. (1993) 2:1541-1545
 #title Differential 3' polyadenylation of the Huntington disease gene results in two mRNA species with variable tissue expression.
 #cross-references MUID:94093536

#accession I54337

#status preliminary
 #molecule_type mRNA
 #residues 2563-3144 #label RES
 #cross-references GB:L20431; NID:g398028; PID:g398029

GENETICS
 #gene GDB:HD
 #cross-references GDB:119307; OMIM:143100
 #map_position 4p16.3-4p16.3
 SUMMARY #length 3144 #molecular-weight 347896 #checksum 1801

Query Match 68.2%; Score 45; DB 2; Length 3144;
 Best Local Similarity 60.0%; Pred. No. 2.93e+01;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2343 EEEEDVPNT 2352
 ||:||||:
 QY 1 EEOQEVPPDT 10

RESULT 13
 ENTRY S37431 #type complete
 TITLE ankyrin 2, neuronal long splice form - human
 ALTERNATE_NAMES ankyrin B, 440K splice form; ankyrin-B; brain ankyrin; non-erythroid ankyrin
 CONTAINS ankyrin 2, short form

```

ORGANISM      #formal_name Homo sapiens #common_name man
DATE          06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
ACCESSIONS    S37431; A39643; B39643; A40334; A49462; S14533; S14569
REFERENCE      Chan, W.
#authors      Chan, W.
#submission   submitted to the EMBL Data Library, September 1993
#accession    S37431
#status       preliminary
#molecule_type mRNA
#residues     1-3924 #label CHA
#cross-references EMBL:Z26634; NID:g406287; PID:g406288
REFERENCE      Otto, E.; Kunimoto, M.; McLaughlin, T.; Bennett, V.
#authors      J. Cell Biol. (1993) 114:241-253
#journal      Isolation and characterization of cDNAs encoding human brain
#title        ankryrins reveal a family of alternatively spliced genes.
#cross-references MUID:91302466
#accession    A39643
#status       preliminary
#molecule_type mRNA
#residues     1-2077 #label OT1
#cross-references GB:X56957
#accession    B39643
#status       preliminary
#molecule_type mRNA
#residues     1-1443,3585-3924 #label OTT
#cross-references EMBL:X56958
REFERENCE      Tse, W.T.; Menninger, J.C.; Yang-Feng, T.L.; Francke, U.;
#authors      Sahr, K.E.; Lux, S.E.; Ward, D.C.; Forget, B.G.
#journal      Genomics (1991) 10:858-866
#title        Isolation and chromosomal localization of a novel
#              nonerythroid ankryrin gene.
#cross-references MUID:9200921
#accession    A40334
#molecule_type DNA
#residues     463-474,'PB',477-495 #label TSE
#cross-references GB:M37123; NID:g178647; PID:g178648
REFERENCE      Chan, W.; Kordeli, E.; Bennett, V.
#authors      J. Cell Biol. (1993) 123:1463-1473
#journal      440-kD ankryrinB: structure of the major developmentally
#title        regulated domain and selective localization in unmyelinated
#              axons.
#cross-references MUID:94075409
#accession    A49462
#status       preliminary; nucleic acid sequence not shown
#molecule_type mRNA
#residues     1-3924 #label RES
#cross-references EMBL:Z26634; NID:g406287; PID:g406288
GENETICS
#gene         GDB:ANK2
#map_position 4q25-4q27
#classification superfamily ankryrin; ankryrin repeat homology
#keywords     alternative splicing
FEATURE
2-3924        #product ankryrin 2, long form #status predicted #label
              MAT\
              MA2\
2-1443,3585-3924 #product ankryrin 2, short form #status predicted #label
              MA2\
63-95         #domain ankryrin repeat homology #label AN01\
98-128        #domain ankryrin repeat homology #label AN02\
129-161       #domain ankryrin repeat homology #label AN03\
162-190       #domain ankryrin repeat homology #label AN04\
191-223       #domain ankryrin repeat homology #label AN05\
232-264       #domain ankryrin repeat homology #label AN06\
265-297       #domain ankryrin repeat homology #label AN07\
298-330       #domain ankryrin repeat homology #label AN08\
331-363       #domain ankryrin repeat homology #label AN09\
364-396       #domain ankryrin repeat homology #label AN10\
397-429       #domain ankryrin repeat homology #label AN11\

```

```

430-462       #domain ankryrin repeat homology #label AN12\
463-495       #domain ankryrin repeat homology #label AN13\
496-528       #domain ankryrin repeat homology #label AN14\
529-561       #domain ankryrin repeat homology #label AN15\
562-594       #domain ankryrin repeat homology #label AN16\
595-627       #domain ankryrin repeat homology #label AN17\
628-660       #domain ankryrin repeat homology #label AN18\
661-693       #domain ankryrin repeat homology #label AN19\
727-759       #domain ankryrin repeat homology #label AN20\
760-792       #domain ankryrin repeat homology #label AN21\
793-825       #domain ankryrin repeat homology #label AN22\
              #domain ankryrin repeat homology #label AN23\
SUMMARY       #length 3924 #molecular-weight 430340 #checksum 3664
Query Match   68.2%; Score 45; DB 2; Length 3924;
Best Local Similarity 40.0%; Pred. No. 2.93e+01;
Matches       4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
Db            3839 DMPPEIPPET 3848
QY            1 EEQEQEVPPDT 10
              ::|::|::|
RESULT        14
ENTRY         S37862
TITLE         hypothetical protein YKL041w - yeast (Saccharomyces
              cerevisiae)
ALTERNATE_NAMES hypothetical protein YKL254
ORGANISM       formal_name Saccharomyces cerevisiae
DATE           03-May-1994 #sequence_revision 03-May-1994 #text_change
ACCESSIONS     S37862; S40650
REFERENCE       S37851
#authors       Purnelle, B.; Skala, J.; van Dyck, L.; Tettelin, H.; Goffeau,
              A.
#submission    submitted to the Protein Sequence Database, March 1994
#accession     S37862
#molecule_type DNA
#residues      1-224 #label PUR
#cross-references EMBL:Z28041; NID:g486051; PID:g486052; MIPS:YKL041w
#experimental_source strain S288C
REFERENCE       S40650
#authors       Purnelle, B.; Tettelin, H.; van Dyck, L.; Skala, J.; Goffeau,
              A.
#journal       Yeast (1993) 9:1379-1384
#title         The sequence of a 17.5 kb DNA fragment on the left arm of
              yeast chromosome XI identifies the protein kinase gene
              ELM1, the DNA primase gene PRI2, a new gene encoding a
              putative histone and seven new open reading frames.
#accession     S40650
#status        nucleic acid sequence not shown
#molecule_type DNA
#residues      1-224 #label PU2
#cross-references EMBL:X71621; NID:g1478265; PID:g666099
#experimental_source strain S288C
GENETICS
#gene         SGD:VPS24
#map_position 11L
#cross-references SGD:S0001524; MIPS:YKL041w
SUMMARY        #length 224 #molecular-weight 26242 #checksum 5062
Query Match    66.7%; Score 44; DB 2; Length 224;
Best Local Similarity 55.6%; Pred. No. 4.48e+01;
Matches        5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db            194 EEQEQEIPDE 202
QY            1 EEQEQEVPPD 9
              ::|::|::|
RESULT        15
ENTRY         H65206
TITLE         uroporphyrinogen decarboxylase (EC 4.1.1.37) - Escherichia

```

```

coli
#formal name Escherichia coli
12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
05-Dec-1997
ACCESSIONS H65206; JN0894; JS0708
REFERENCE A64720
#authors
  Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
  Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
  Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
  Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
  Y.
#journal
  Science (1997) 277:1453-1462
#title
  The complete genome sequence of Escherichia coli K-12.
#cross-references MIM:97426617
#accession H65206
  ##status nucleic acid sequence not shown; translation not shown
  ##molecule_type DNA
  ##residues 1-354 #label BLAT
  ##cross-references GB:A5000473; GB:U00096; NID:g2367336; PID:g2367337;
    UWGP:b3997
  ##experimental_source strain K-12, substrain MG1655
REFERENCE JN0894
#authors
  Nishimura, K.; Nakayashiki, T.; Inokuchi, H.
#journal
  Gene (1993) 133:109-113
#title
  Cloning and sequencing of the hemE gene encoding
  uroporphyrinogen III decarboxylase (UPD) from Escherichia
  coli K-12.
#accession JN0894
  ##status nucleic acid sequence not shown
  ##molecule_type DNA
  ##residues 1-71,'RSS','75','R','Y','79-82','I','84-88','SSILKP','95','KV',
    98,'VLPRQI','104-250','SAIV','A','257-354 #label NI2
  ##cross-references GB:D12624; NID:g216563; PID:d1002638; PID:g216564
  ##experimental_source strain K-12
COMMENT
  This enzyme catalyzes the decarboxylation of uroporphyrinogen III
  to coproporphyrinogen.
GENETICS
#gene hemeF
#map_position 90 min
CLASSIFICATION #superfamily uroporphyrinogen decarboxylase
KEYWORDS carbon-carbon lyase; carboxy-lyase; lyase; porphyrin
  biosynthesis
SUMMARY #length 354 #molecular-weight 39248 #checksum 7044
Query Match 66.7%; Score 44; DB 2; Length 354;
Best Local Similarity 57.1%; Pred. NO. 4.48e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 330 HODVPPPE 336
QY :|:|:|:
  3 QQEVPPD 9

```

Search completed: Thu May 13 15:32:28 1999
 Job time : 14 secs.

W A P E R L E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu May 13 15:32:47 1999; MasPar time 2.01 Seconds
Tabular output not generated. 133.387 Million cell updates/sec

Title: >US-09-040-485-7
Description: (1-10) from US09040485.pap
Perfect Score: 66
Sequence: 1 EEQQEVPDPDT 10
Scoring table: PAM 150
Gap 15
Searched: 74019 seqs, 26840295 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: swiss-prot35
1:swissprot
Statistics: Mean 22.125; Variance 25.762; scale 0.859

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	65	98.5	754	1 ASPH_BOVIN	ASPARTYL/ASPARAGINYL B	2.67e-04
2	65	98.5	757	1 ASPH_HUMAN	ASPARTYL/ASPARAGINYL B	2.67e-04
3	55	83.3	816	1 YM35_YEAST	HYPOTHETICAL 95.1 KD P	7.70e-02
4	49	74.2	226	1 PRL_MESAU	PROLACTIN PRECURSOR (P	1.79e+00
5	48	72.7	815	1 LU15_HUMAN	PUTATIVE TUMOR SUPPRES	2.95e+00
6	46	69.7	167	1 YJG5_YEAST	HYPOTHETICAL 18.8 KD P	7.87e+00
7	46	69.7	775	1 LY54_EMENI	HOMOACONITASE PRECURS	7.87e+00
8	45	68.2	217	1 U119_CAEBR	UNC-119 PROTEIN.	1.27e+01
9	45	68.2	671	1 PEXH_YARLI	PEROXISOMAL MEMBRANE P	1.27e+01
10	45	68.2	852	1 SRCH_RABIT	SARCOPLASMIC RETICULUM	1.27e+01
11	45	68.2	1839	1 ANKC_HUMAN	ANKYRIN, BRAIN VARIANT	1.27e+01
12	45	68.2	3144	1 HD_HUMAN	HUNTINGTIN (HUNTINGTON	1.27e+01
13	45	68.2	3924	1 ANKB_HUMAN	ANKYRIN, BRAIN VARIANT	1.27e+01
14	44	66.7	224	1 YKEL_YEAST	HYPOTHETICAL 26.2 KD P	2.04e+01
15	44	66.7	354	1 DCUP_ECOLI	UROPORPHYRINOGEN DECAR	2.04e+01
16	44	66.7	414	1 NOP3_HUMAN	NUCLEOLAR PROTEIN 3 (M	2.04e+01
17	44	66.7	700	1 MYBB_HUMAN	MYB-RELATED PROTEIN B	2.04e+01
18	44	66.7	755	1 RREL_HUMAN	RAS-RESPONSIVE ELEMENT	2.04e+01
19	43	65.2	140	1 ANF_CHICK	ATRIAL NATRIURETIC FAC	3.24e+01
20	43	65.2	184	1 K501_ACTCH	FRUIT PROTEIN PKM1501	3.24e+01
21	43	65.2	395	1 CG2A_CHICK	G2/MITOTIC-SPECIFIC CY	3.24e+01
22	43	65.2	404	1 ARRS_BOVIN	S-ARRESTIN (RETINAL S-	3.24e+01
23	43	65.2	405	1 ARRS_HUMAN	S-ARRESTIN (RETINAL S-	3.24e+01

ALIGNMENTS

RESULT	1	ASPH_BOVIN	STANDARD;	PRT;	754 AA.
ID	AC	Q28056;			
DT	01-NOV-1997	(REL. 35, CREATED)			
DT	01-NOV-1997	(REL. 35, LAST SEQUENCE UPDATE)			
DT	01-NOV-1997	(REL. 35, LAST ANNOTATION UPDATE)			
DE	ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-HYDROXYLASE) (ASP BETA-HYDROXYLASE) (PEPTIDE-ASPARTATE BETA-HYDROXYLASE)				
GN	ASPH.				
OS	BOS TAURUS (BOVINE).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;				
OC	EUTHERIA; ARTIODACTYLA.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=LIVER, AND BRAIN;				
RX	MEDLINE: 92332546.				
RA	JIA S., VANDUSEN W.J., DIEHL R.E., KOHL N.E., DIXON R.A.F., ELLISTON K.O., STERN A.M., FRIEDMAN P.A.; J. BIOL. CHEM. 267:14322-14327(1992).				
RN	[2]				
RP	SEQUENCE OF 289-385 AND 615-641.				
RC	TISSUE=LIVER;				
RX	MEDLINE: 91310689.				
RA	WANG Q., VANDUSEN W.J., PETROSKI C.J., GARSKY V.M., STERN A.M., FRIEDMAN P.A.; J. BIOL. CHEM. 266:14004-14010(1991).				
CC	-1- FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF PROTEINS.				
CC	-1- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) - PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).				
CC	-1- COFACTOR: IRON.				
CC	-1- SUBUNIT: MONOMER.				
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM.				
CC	-1- PTM: MIGHT BE PROCESSED TO THE 56 KD (AA 289-754) OR 52 KD (AA 311-754) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM.				
EMBL	M91213; G162694;				
DR	OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR; ENDOPLASMIC RETICULUM.				
KW	DOMAIN 1 57				
FT	DOMAIN 58 78				
FT	TRANSMEM				
FT	DOMAIN 79 754				
FT	DOMAIN 9 12				
FT	DOMAIN 14 21				
FT	DOMAIN				

```
FT DOMAIN 318 328 POLY-LYS.
FT CARBOHYD 96 96 POTENTIAL.
FT CARBOHYD 466 466 POTENTIAL.
FT CARBOHYD 702 702 POTENTIAL.
SQ SEQUENCE 754 AA; 84998 MW; 608861B2 CRC32;

Query Match 98.5%; Score 65; DB 1; Length 754;
Best Local Similarity 90.0%; Pred. No. 2.67e+04;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 300 EEOQEVPPET 309
QY 1 EEOQEVPPD 10

RESULT 2
ID ASPH_HUMAN STANDARD; PRT; 757 AA.
AC Q12797;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.4.11.16) (ASPARTATE BETA-
DE HYDROXYLASE) (ASP BETA-HYDROXYLASE) (PEPTIDE-ASPARTATE BETA-
DE DIOXYGENASE).
GN ASPH.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95121937.
RA KORIOTH F., GIEFFERS C., FREY J.;
RL GENE 150:395-399(1994).
CC -!- FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN
CC CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF
CC PROTEINS.
CC -!- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) -
CC PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).
CC -!- COFACTOR: IRON.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC
CC RETICULUM.
CC -!- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES TESTED.
CC -!- PTM: MIGHT BE PROCESSED TO THE 56 KD (AA 274-757) OR 52 KD (AA
CC 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC EMBL; 003109; E82591; -.
DR MIM; 600582; -.
KW OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR;
KW ENDOPLASMIC RETICULUM.
FT DOMAIN 1 54 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 55 75 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 76 757 LUMENAL (POTENTIAL).
FT DOMAIN 13 20 POLY-SER.
FT DOMAIN 323 332 POLY-LYS.
FT CARBOHYD 452 452 POTENTIAL.
FT CARBOHYD 705 705 POTENTIAL.
SQ SEQUENCE 757 AA; 85498 MW; AE6AFC24 CRC32;

Query Match 98.5%; Score 65; DB 1; Length 757;
Best Local Similarity 90.0%; Pred. No. 2.67e+04;
Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 304 EEOQEVPPET 313
QY 1 EEOQEVPPD 10

RESULT 3
ID YN35_YEAST STANDARD; PRT; 816 AA.
AC Q03823;
DT 01-NOV-1997 (REL. 35, CREATED)
```

```
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 95.1 KD PROTEIN IN IMPI-HLJ1 INTERGENIC REGION.
GN YN160W OR YN8520.09.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA HUNT S., BOWMAN S., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; Z49705; G825565; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 816 AA; 95096 MW; A897E595 CRC32;

Query Match 83.3%; Score 55; DB 1; Length 816;
Best Local Similarity 77.8%; Pred. No. 7.70e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 24 EDOQEVPPQ 32
QY 1 EEOQEVPPD 9

RESULT 4
ID PRL_MESAU STANDARD; PRT; 226 AA.
AC P37884;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE PROLACTIN PRECURSOR (PRL).
GN PRL.
OS MESOCRICETUS AURATUS (GOLDEN HAMSTER).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92063850.
RA SOUTHWARD J.N., SANCHEZ-JIMENEZ F., CAMPBELL G.T., TALAMANTES F.;
RL ENDOCRINOLOGY 129:2965-2971(1991).
CC -!- FUNCTION: PROLACTIN ACTS PRIMARILY ON THE MAMMARY GLAND BY
CC PROMOTING LACTATION.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
DR EMBL; S86296; G239354; -.
DR PIR; A49159; A49159.
DR PROSITE; PS00266; SOMATOTROPIN_1; 1.
DR PROSITE; PS00338; SOMATOTROPIN_2; 1.
KW HORMONE; PARTURITION; LACTATION; PITUITARY; SIGNAL.
FT CHAIN 1 29 BY SIMILARITY.
FT SIGNAL 30 226 PROLACTIN.
FT DISULFID 33 38 BY SIMILARITY.
FT DISULFID 85 201 BY SIMILARITY.
FT DISULFID 218 226 BY SIMILARITY.
SQ SEQUENCE 226 AA; 25582 MW; D5A4C5BC CRC32;

Query Match 74.2%; Score 49; DB 1; Length 226;
Best Local Similarity 66.7%; Pred. No. 1.79e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 97 EEAQQVPPPE 105
QY 1 EEOQEVPPD 9

RESULT 5
ID LU15_HUMAN STANDARD; PRT; 815 AA.
AC P52756;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE PUTATIVE TUMOR SUPPRESSOR LOC1A15.
GN LOC1A15.
```

OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 CC EUTHERIA; PRIMATES.
 RN SEQUENCE FROM N.A.
 RA BADER S., LATIF F., DUH F., WEI M., KASHUBA V., SEKIDO Y., LEE C.,
 RA KOONIN E., ZARAFESKY E., KLEIN G., MINNA J.D., LERMAN M.;
 RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: CONTAINS 2 RNA RECOGNITION MOTIFS (RNP).
 CC -1- SIMILARITY: HIGH, TO HUMAN DXS8237E.
 DR EMBL: U23946; G1244404; -.
 DR PROSITE: PS00030; RNP_1; 2.
 KW ANTI-ONCOGENE: RNA-BINDING; NUCLEAR PROTEIN.
 FT DOMAIN 140 147 RNA-BINDING (RNP1) (BY SIMILARITY).
 FT DOMAIN 274 281 RNA-BINDING (RNP1) (BY SIMILARITY).
 SQ SEQUENCE 815 AA; 92073 MW; 30EF5EB2 CRC32;

Query Match 72.7%; Score 48; DB 1; Length 815;
 Best Local Similarity 60.0%; Pred. No. 2.95e+00;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 210 DSEQVPPGT 219

Qy 1 EEQEVPPDT 10

RESULT 6
 ID YJG5_YEAST STANDARD; PRT; 167 AA.
 AC P40366;

DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 18.8 KD PROTEIN IN SMC3-MRPL8 INTERGENIC REGION.
 GN YJL065C OR J1115 OR HRD167.
 OS SACHAROMYCES CEREVISIAE (BAKER'S YEAST).
 CC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE; 95282514.

RA VANDENBOL M., DURAND P., DION C., PORTETELLE D., HILGER F.;
 RL YEAST 11:57-60(1995).

CC -1- SIMILARITY: TO YEAST DNA POLYMERASE EPSILON, SUBUNIT C (DPB3).
 DR EMBL: Z34288; G499003; -.
 DR EMBL: Z49340; G1008213; -.
 DR PIR: S47127; S47127.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 167 AA; 18792 MW; B5DC5E3C CRC32;

Query Match 69.7%; Score 46; DB 1; Length 167;
 Best Local Similarity 62.5%; Pred. No. 7.87e+00;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 68 OQQQVPP 75

Qy 1 EEQEVPP 8

RESULT 7
 ID LYS4_EMENI STANDARD; PRT; 775 AA.
 AC O92412;

DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE HOMOACONITASE PRECURSOR (EC 4.2.1.36) (HOMOACONITATE HYDRATASE).
 GN LYSF.

OS EMERICELLA NIDULANS (ASPERGILLUS NIDULANS).

CC EUKARYOTA; FUNGI; ASCOMYCOTINA; PLECTOMYCETES; EUROTIALES.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 97411901.

RA WEIDNER G., STEFFAN B., BRAKHAGE A.A.;

RL MOL. GEN. GENET. 255:237-247(1997).
 CC -1- FUNCTION: RESPONSIBLE FOR THE DEHYDRATION OF CIS-HOMOACONITATE TO
 CC HOMOISOCITRIC ACID.
 CC -1- CATALYTIC ACTIVITY: 2-HYDROXYBUTANE-1,2,4-TRICARBOXYLATE - BUT-1-
 CC ENE-1,2,4-TRICARBOXYLATE + H(2)O.
 CC -1- PATHWAY: THIRD STEP IN LYSINE BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
 DR EMBL: X99624; E265392; -.
 DR PROSITE: PS00450; ACONITASE_1; 1.
 DR PROSITE: PS01244; ACONITASE_2; 1.
 KW LYSINE BIOSYNTHESIS; LYASE; MITOCHONDRION; TRANSIT PEPTIDE;
 KW IRON-SULFUR.
 FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
 FT CHAIN ? 775 HOMOACONITASE.
 FT METAL 394 394 IRON (IRON-SULFUR CLUSTER)
 FT METAL 462 462 (BY SIMILARITY).
 FT METAL 462 462 IRON (IRON-SULFUR CLUSTER)
 FT METAL 465 465 (BY SIMILARITY).
 FT METAL 465 465 IRON (IRON-SULFUR CLUSTER)
 FT METAL 465 465 (BY SIMILARITY).
 SQ SEQUENCE 775 AA; 84038 MW; 86EDE640 CRC32;

Query Match 69.7%; Score 46; DB 1; Length 775;

Best Local Similarity 62.5%; Pred. No. 7.87e+00;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 619 QDDVPPPT 626

Qy 3 QQEVPPDT 10

RESULT 8

ID U119_CABER STANDARD; PRT; 217 AA.

AC Q17297; 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE UNC-119 PROTEIN.

GN UNC-119

OS CAENORHABDITIS BRIGGSAE.

OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
 RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 97149282.

RA MADURO M.F., PILGRIM D.B.;

RL GENE 183:77-85(1996).

CC -1- FUNCTION: REQUIRED FOR THE ESTABLISHMENT OR FUNCTION OF THE

CC NERVOUS SYSTEM (BY SIMILARITY).

CC -1- SIMILARITY: STRONG, TO MAMMALIAN RETINAL PROTEIN 4; WEAK, TO

CC C.ELEGANS C27H5.1.

DR EMBL: U45326; G1181703; -.

KW DEVELOPMENTAL PROTEIN.

SQ SEQUENCE 217 AA; 25094 MW; F477F555 CRC32;

Query Match 68.2%; Score 45; DB 1; Length 217;

Best Local Similarity 50.0%; Pred. No. 1.27e+01;

Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 4 EQQSIPPGS 13

Qy 1 EEQEVPPDT 10

RESULT 9

ID PEXH_YARLI STANDARD; PRT; 671 AA.

AC P87200;

DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE PEROXISOMAL MEMBRANE PROTEIN PEX17 (PEROXIN-17).

GN PEX17

OS YARROWIA LIPOLYTICA (CANDIDA LIPOLYTICA).

OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.

RN SEQUENCE FROM N.A.
 RP STRAIN=E122;
 RC MEDLINE: 97265383.
 RA SMITH J.J., SZILARD R.K., MARELLI M., RACHUBINSKI R.A.;
 RL MOL. CELL. BIOL. 17:2511-2520(1997).
 CC -!- FUNCTION: INVOLVED IN PEROXISOME BIOSYNTHESIS. REQUIRED FOR THE
 CC IMPORT OF A SUBSET OF MATRIX PROTEINS.
 CC -!- SUBCELLULAR LOCATION: PEROXISOMAL MEMBRANE-ASSOCIATED.
 DR EMBL: U73028; G2039277;
 KW PEROXISOME; TRANSMEMBRANE;
 FT TRANSMEM 127 147 .POTENTIAL.
 FT TRANSMEM 187 207 .POTENTIAL.
 FT TRANSMEM 212 232 .POTENTIAL.
 FT TRANSMEM 258 278 .POTENTIAL.
 FT TRANSMEM 318 338 .POTENTIAL.
 FT TRANSMEM 372 392 .POTENTIAL.
 FT TRANSMEM 475 495 .POTENTIAL.
 FT TRANSMEM 504 524 .POTENTIAL.
 SQ SEQUENCE 671 AA; 75588 MW; BC5618B9 CRC32;

Query Match 68.2%; Score 45; DB 1; Length 671;
 Best Local Similarity 50.0%; Pred. No. 1.27e+01;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 573 OEQEDIPP 580
 QY :||:||||
 1 EEQEVPP 8

RESULT 10
 ID SRCH RABIT STANDARD; PRT; 852 AA.
 AC P16230;
 DT 01-APR-1990 (REL. 14, CREATED)
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
 DE SARCOPLASMIC RETICULUM HISTIDINE-RICH CALCIUM-BINDING PROTEIN
 DE PRECURSOR (HCP).
 OS EUCALYPTUS GULONULUS (RABBIT).
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; LAGOMORPHA.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-SKELETAL MUSCLE;
 RX MEDLINE: 90036884.
 RA HOFMANN S.L., GOLDBSTEIN J.L., ORTH K., MOOMAW C.R., SLAUGHTER C.A.,
 RA BROWN M.S.;
 RL J. BIOL. CHEM. 264:18083-18090(1989).
 CC -!- FUNCTION: HCP MAY PLAY A ROLE IN THE REGULATION OF CA(2+)
 CC SEQUESTRATION OR RELEASE IN THE SR OF SKELETAL AND CARDIAC
 CC MUSCLE.
 CC -!- SUBCELLULAR LOCATION: SARCOPLASMIC RETICULUM LUMEN.
 CC -!- SIMILARITY: STRONG TO HUMAN HRC.
 DR EMBL: J05080; G165100; --
 DR PIR: A34373; A34373.
 DR PROSITE: PS00328; HCP; 10.
 KW CALCIUM-BINDING; SIGNAL; REPEAT.
 FT SIGNAL 1 27 .POTENTIAL.
 FT CHAIN 28 852 HISTIDINE-RICH CALCIUM-BINDING PROTEIN.
 FT MOD_RES 28 28 BLOCKED.
 FT DOMAIN 59 100 2 X APPROXIMATE TANDEM REPEATS.
 FT REPEAT 59 79 1-1.
 FT REPEAT 80 100 1-2.
 FT DOMAIN 199 470 10 X TANDEM REPEATS, ACIDIC.
 FT REPEAT 199 224 2-1.
 FT REPEAT 225 233 2-2.
 FT REPEAT 254 282 2-3.
 FT REPEAT 283 310 2-4.
 FT REPEAT 311 339 2-5.
 FT REPEAT 340 367 2-6.
 FT REPEAT 368 395 2-7.
 FT REPEAT 396 423 2-8.
 FT REPEAT 424 451 2-9.

FT REPEAT 452 470 2-10.
 FT DOMAIN 471 585 4 X APPROXIMATE TANDEM REPEATS.
 FT DOMAIN 721 733 POLY-GLU.
 FT DOMAIN 780 826 METAL-BINDING (POTENTIAL).
 SQ SEQUENCE 852 AA; 96117 MW; 3C1DF781 CRC32;
 Query Match 68.2%; Score 45; DB 1; Length 852;
 Best Local Similarity 55.6%; Pred. No. 1.27e+01;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 Db 190 EEEEVSP 198
 QY ||:||||
 1 EEQEVPP 9
 RESULT 11
 ID ANKC HUMAN STANDARD; PRT; 1839 AA.
 AC Q01485;
 DT 01-APR-1993 (REL. 25, CREATED)
 DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE ANKYRIN, BRAIN VARIANT 2 (ANKYRIN B) (ANKYRIN, NONERYTHROID)
 DE (FRAGMENT).
 GN ANK2.
 OS HOMO SAPIENS (HUMAN).
 OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; PRIMATES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN STEM;
 RX MEDLINE: 91302466.
 RA OTTO E., KUNIMOTO M., MCLAUGHLIN T., BENNETT V.;
 RL J. CELL BIOL. 114:241-253(1991).
 RN [2]
 RP SEQUENCE OF 463-495 FROM N.A.
 RX MEDLINE: 92009921.
 RA TSE W.T., MENNINGER J.C., YANG-FENG T.L., FRANCKE U., SAHR K.E.,
 RA LUX S.E., WARD D.C., FORGET B.G.;
 RA GENOMICS 10:858-866(1991).
 CC -!- FUNCTION: ANKYRINS ATTACH INTEGRAL MEMBRANE PROTEINS TO CYTO-
 CC SKELETAL ELEMENTS; THEY BIND TO THE ERYTHROCYTE MEMBRANE PROTEIN
 CC BAND 4.2, TO NA-K ATPASE, TO THE LYMPHOCYTE MEMBRANE PROTEIN GP85,
 CC AND TO THE CYTOSKELETAL PROTEINS FODRIN, TUBULIN, VIMENTIN AND
 CC DESMIN. ERYTHROCYTE ANKYRINS ALSO LINK SPECTRIN (BETA CHAIN) TO
 CC THE CYTOPLASMIC DOMAIN OF THE ERYTHROCYTES ANION EXCHANGE PROTEIN;
 CC THEY RETAIN MOST OR ALL OF THESE BINDING FUNCTIONS.
 CC -!- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL
 CC CELLS THROUGHOUT THE BRAIN.
 CC -!- ALTERNATIVE PRODUCTS: THE TWO BRAIN VARIANTS ARE PRODUCED BY
 CC ALTERNATIVE SPLICING OF GENE ANK2.
 CC -!- SIMILARITY: CONTAINS 24 ANK REPEATS.
 DR EMBL: X56958; G29491; --
 DR EMBL: M37123; G178648; --
 DR PIR: S14569; S14569.
 DR PIR: B39643; B39643.
 DR MIM: 106410; --
 DR PROSITE: PS50017; DEATH DOMAIN: 1.
 KW CYTOSKELETON; ALTERNATIVE SPLICING; REPEAT; ANK REPEAT;
 MULTIGENE FAMILY.
 FT DOMAIN 63 792 22 X ANK MOTIF REPEATS.
 FT REPEAT 63 95 ANK MOTIF 1.
 FT REPEAT 96 128 ANK MOTIF 2.
 FT REPEAT 129 161 ANK MOTIF 3.
 FT REPEAT 162 190 ANK MOTIF 4.
 FT REPEAT 191 231 ANK MOTIF 5.
 FT REPEAT 232 264 ANK MOTIF 6.
 FT REPEAT 265 297 ANK MOTIF 7.
 FT REPEAT 298 330 ANK MOTIF 8.
 FT REPEAT 331 363 ANK MOTIF 9.
 FT REPEAT 364 396 ANK MOTIF 10.
 FT REPEAT 397 429 ANK MOTIF 11.
 FT REPEAT 430 462 ANK MOTIF 12.
 FT REPEAT 463 495 ANK MOTIF 13.

RP SEQUENCE OF 1212-1290 FROM N.A.
RA MUNGALL A., ODELL C.;
RN SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
[7]
RP SEQUENCE OF 1291-1860 FROM N.A.
RA MUNGALL A.;
RN SUBMITTED (APR-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
[8]
RP SEQUENCE OF 1862-2820 FROM N.A.
RA BUCK D.;
RN SUBMITTED (MAY-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
[9]
RP SEQUENCE OF 2563-3144 FROM N.A.
RN TISSUE=FRONTAL CORTEX, BRAIN, RETINA, CAUDATE, AND MUSCLE;
RC MEDLINE: 94093536.
RA LIN B., ROMMENS J.M., GRAHAM R.K., KALCHMAN M., MACDONALD H.,
RA NASIR J., DELANEY A., GOLDBERG J.P., HAYDEN M.R.;
RA HUM. MOL. GENET. 2:1541-1545(1993).
[10]
RP SUBCELLULAR LOCATION.
RN MEDLINE: 95375771.
RX TROTIER Y., DEVYS D., IMBERT G., SAUDOU F., AN I., LUTZ Y., WEBER C.,
RA ACID Y., HIRSCH E.C., MANDEL J.-L.;
RN NAT. GENET. 10:104-110(1995).
[11]
RN CLEAVAGE BY APOPAIN.
RX MEDLINE: 96331285.
RA GOLDBERG Y.P., NICHOLSON D.W., RASPER D.M., KALCHMAN M.A., KOIDE H.B.,
RA GRAHAM R.K., BROMM M., KAZEMI-ESFARJANI P., THORNBERRY N.A.,
RA VAILLANCOURT J.P., HAYDEN M.R.;
RN NAT. GENET. 13:442-449(1996).
CC -1- FUNCTION: MAY PLAY A ROLE IN MICROTUBULE-MEDIATED TRANSPORT OR
CC VESICLE FUNCTION
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN A VARIETY OF TISSUES WITH
CC THE HIGHEST LEVEL EXPRESSION IN THE BRAIN (NERVE FIBRES,
CC VARICOSITIES, AND NERVE ENDINGS). IN THE BRAIN, THE REGIONS WHERE
CC IT CAN BE MAINLY FOUND ARE THE CEREBELLAR CORTEX, THE NEOCORTEX,
CC THE STRIATUM, AND THE HIPPOCAMPAL FORMATION.
CC -1- PTM: CLEAVED BY APOPAIN DOWNSTREAM OF THE POLYGLUTAMINE STRETCH.
CC THE RESULTING AMINO-TERMINAL FRAGMENT IS CYTOTOXIC AND PROVOKES
CC APOPTOSIS.
CC -1- POLYMORPHISM: THE POLY-GLN REGION OF HD IS HIGHLY POLYMORPHIC (10
CC TO 35 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO ABOUT
CC 36-120 REPEATS IN HD PATIENTS. THE REPEAT LENGTH USUALLY INCREASES
CC IN SUCCESSIVE GENERATIONS, BUT CONTRACTS ALSO ON OCCASION. THE
CC LONGER EXPANSIONS RESULT IN EARLIER ONSET AND MORE SEVERE CLINICAL
CC MANIFESTATIONS OF THE DISEASE. THE ADJACENT POLY-PRO REGION IS
CC ALSO POLYMORPHIC AND VARIES BETWEEN 7-12 RESIDUES. POLYGLUTAMINE
CC EXPANSION LEADS TO ELEVATED SUSCEPTIBILITY TO APOPAIN CLEAVAGE AND
CC LIKELY RESULT IN ACCELERATED NEURONAL APOPTOSIS.
CC -1- DISEASE: DEFECTS IN HD ARE THE CAUSE OF HUNTINGTON'S DISEASE, AN
CC AUTOSOMAL DOMINANT NEURODEGENERATIVE DISORDER CHARACTERIZED BY
CC INVOLUNTARY MOVEMENTS (CHOREA), GENERAL MOTOR IMPAIRMENT,
CC PSYCHIATRIC DISORDERS AND DEMENTIA. ONSET OF THE DISEASE OCCURS
CC USUALLY IN THE THIRD OR FOURTH DECADE OF LIFE AND SYMPTOMS
CC PROGRESSIVELY WORSEN LEADING TO DEATH IN 10 TO 20 YEARS. IT
CC AFFECTS 1 IN 10,000 INDIVIDUALS OF EUROPEAN ORIGIN. NEUROPATHOLOGY
CC OF HUNTINGTON'S DISEASE DISPLAYS A DISTINCTIVE PATTERN WITH LOSS
CC OF NEURONS, SPECIALLY IN THE CAUDATE AND PUTAMEN (STRIATUM).
CC -1- SIMILARITY: STRONGLY CONSERVED BETWEEN HUMAN, RAT AND MOUSE.
CC -1- DATABASE: NAME=HotMolcbase; NOTE=HD entry;
CC WWW="http://bioinformatics.welzmann.ac.il/hotmolecbase/entries/hun/
CC EMBL: L12392; G454415; -;
DR EMBL: L34020; -; NOT_ANNOTATED_CDS.
DR EMBL: L27350; -; NOT_ANNOTATED_CDS.
DR EMBL: L27351; -; NOT_ANNOTATED_CDS.
DR EMBL: L27352; -; NOT_ANNOTATED_CDS.
DR EMBL: L27353; -; NOT_ANNOTATED_CDS.
DR EMBL: L27354; -; NOT_ANNOTATED_CDS.
DR EMBL: L268756; E218669; -;
DR EMBL: L249155; G794068; -;
DR EMBL: L249208; G798838; -;

DR EMBL; 269649; E223397; -
DR EMBL; 249154; G794066; -
DR EMBL; 249769; G840782; -
DR EMBL; L20431; G398029; -
DR MIM; 143100; -
KW DISEASE MUTATION; POLYMORPHISM; TRIPLET REPEAT EXPANSION; APOPTOSIS.
FT DOMAIN 18 40 POLY-GLN.
FT DOMAIN 41 51 POLY-PRO.
FT DOMAIN 65 80 POLY-PRO.
FT DOMAIN 1439 1442 POLY-THR.
FT DOMAIN 2343 2347 POLY-GLU.
FT DOMAIN 2640 2645 POLY-GLU.
FT VARIANT 38 40 MISSING.
FT SITE 513 514 CLEAVAGE BY APOPAIN (POTENTIAL).
FT SITE 530 531 CLEAVAGE BY APOPAIN (POTENTIAL).
FT SITE 552 553 CLEAVAGE BY APOPAIN (POTENTIAL).
FT SITE 589 590 CLEAVAGE BY APOPAIN (POTENTIAL).
FT CONFLICT 2788 2788 V -> I (IN REF. 10).
SQ SEQUENCE 3144 AA; 347855 MW; 8BF8AE44 CRC32;

Query Match 68.2%; Score 45; DB 1; Length 3144;
Best Local Similarity 60.0%; Pred. No. 1.27e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 2343 EEEVDPT 2352
QY 1 EEEVDPT 10
|::|||::|

RESULT 13
ID ANKB_HUMAN STANDARD; PRT; 3924 AA.
AC Q01484;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE ANKYRIN, BRAIN VARIANT 1 (ANKYRIN B) (ANKYRIN, NONERYTHROID).
GN ANK2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
[1]
RN SEQUENCE FROM N.A.
RC TISSUE-BRAIN STEM;
RX MEDLINE; 94075409.
RA CHAN W., KORDELI E., BENNETT V.;
RL J. CELL BIOL. 123:1463-1473(1993).
[2]
RN SEQUENCE OF 1-2077 FROM N.A.
RC TISSUE-BRAIN STEM;
RX MEDLINE; 91302466.
RA OTTO E., KUNIMOTO M., MCLAUGHLIN T., BENNETT V.;
RL J. CELL BIOL. 114:241-253(1991).
CC -!- FUNCTION: ANKYRINS ATTACH INTEGRAL MEMBRANE PROTEINS TO CYTO-
SKELETAL ELEMENTS. THEY BIND TO THE ERYTHROCYTE MEMBRANE PROTEIN
BAND 4.2, TO NA-K ATPASE, TO THE LYMPHOCYTE MEMBRANE PROTEIN GP85,
AND TO THE CYTOSKELETAL PROTEINS FODRIN, TUBULIN, VIMENTIN AND
DESMIN. ERYTHROCYTE ANKYRINS ALSO LINK SPECTRIN (BETA CHAIN) TO
THE CYTOPLASMIC DOMAIN OF THE ERYTHROCYTES ANION EXCHANGE PROTEIN;
THEY RETAIN MOST OR ALL OF THESE BINDING FUNCTIONS.
CC -!- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL
CELLS THROUGHOUT THE BRAIN.
CC -!- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE
AND FUNCTION (POTENTIAL).
CC -!- ALTERNATIVE PRODUCTS: THE TWO BRAIN VARIANTS ARE PRODUCED BY
ALTERNATIVE SPLICING OF GENE ANK2.
CC -!- SIMILARITY: CONTAINS 24 ANK REPTS.
DR EMBL; 226634; G406288; -
DR EMBL; X56957; G29489; -
DR PIR; S14533; S14533.
DR PIR; A39643; A39643.
DR MIM; 106410; -
DR PROSITE; PSS0017; DEATH_DOMAIN; 1.

KW CYTOSKELETON; ALTERNATIVE SPLICING; REPEAT; ANK REPEAT;
KW PHOSPHORYLATION; MULTIGENE FAMILY.
FT DOMAIN 63 792 24 X ANK MOTIF REPEATS.
FT REPEAT 63 95 ANK MOTIF 1.
FT REPEAT 96 128 ANK MOTIF 2.
FT REPEAT 129 161 ANK MOTIF 3.
FT REPEAT 162 190 ANK MOTIF 4.
FT REPEAT 191 231 ANK MOTIF 5.
FT REPEAT 232 264 ANK MOTIF 6.
FT REPEAT 265 297 ANK MOTIF 7.
FT REPEAT 298 330 ANK MOTIF 8.
FT REPEAT 331 363 ANK MOTIF 9.
FT REPEAT 364 396 ANK MOTIF 10.
FT REPEAT 397 429 ANK MOTIF 11.
FT REPEAT 430 462 ANK MOTIF 12.
FT REPEAT 463 495 ANK MOTIF 13.
FT REPEAT 496 528 ANK MOTIF 14.
FT REPEAT 529 561 ANK MOTIF 15.
FT REPEAT 562 594 ANK MOTIF 16.
FT REPEAT 595 627 ANK MOTIF 17.
FT REPEAT 628 660 ANK MOTIF 18.
FT REPEAT 661 693 ANK MOTIF 19.
FT REPEAT 694 726 ANK MOTIF 20.
FT REPEAT 727 759 ANK MOTIF 21.
FT REPEAT 760 792 ANK MOTIF 22.
FT DOMAIN 1773 1950 REPEAT-RICH REGION.
FT REPEAT 1773 1784 REPEAT A.
FT REPEAT 1785 1796 REPEAT A.
FT REPEAT 1797 1808 REPEAT A.
FT REPEAT 1809 1820 REPEAT A.
FT REPEAT 1821 1832 REPEAT A.
FT REPEAT 1833 1844 REPEAT A.
FT REPEAT 1845 1856 REPEAT A.
FT REPEAT 1857 1867 REPEAT A (APPROXIMATE).
FT REPEAT 1868 1879 REPEAT A.
FT REPEAT 1880 1891 REPEAT A.
FT REPEAT 1892 1902 REPEAT A (APPROXIMATE).
FT REPEAT 1903 1914 REPEAT A.
FT REPEAT 1915 1926 REPEAT A.
FT REPEAT 1927 1938 REPEAT A.
FT REPEAT 1939 1950 REPEAT A.
FT DOMAIN 3536 3620 DEATH DOMAIN.
SQ SEQUENCE 3924 AA; 430337 MW; 0CCC249D CRC32;

Query Match 68.2%; Score 45; DB 1; Length 3924;
Best Local Similarity 40.0%; Pred. No. 1.27e+01;
Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 3839 DDMPEIPPET 3848
QY 1 EEEVDPT 10
::|::|::|

RESULT 14
ID YKEL_YEAST STANDARD; PRT; 224 AA.
AC P36095;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 26.2 KD PROTEIN IN SPC42-PTM1 INTERGENIC REGION.
GN YKL041W OR YKL254.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-S288C;
RX MEDLINE; 94205268.
RA PURNELLE B., TETTELIN H., VAN DYCK L., SKALA J., GOFFEAU A.;
RL YEAST 9:1379-1384(1993).
DR EMBL; X71621; G666099; -
DR EMBL; Z28041; G486052; -
DR PIR; S37862; S37862.
KW HYPOTHETICAL PROTEIN.

```

SQ SEQUENCE 224 AA; 26242 MW; 3FEAA543 CRC32;
Query Match 66.7%; Score 44; DB 1; Length 224;
Best Local Similarity 55.6%; Pred. No. 2.04e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 194 EEEQEIPDE 202
||:|:|:|
QY 1 EEQEVPPD 9

RESULT 15
ID DCUP_ECOLI STANDARD; PRT; 354 AA.
AC P29680; P78135;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE UROPORPHYRINOGEN DECARBOXYLASE (EC 4.1.1.37) (UPD).
GN HEME.
OS ESCHERICHIA COLI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 94040783.
RA NISHIMURA K., INOKUCHI H.;
RL GENE 133:109-113(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 94089392.
RA BLATTNER F.R., BURLAND V.D., PLUNKETT G. III, SOFIA H.J.,
RA DANIELS D.L.;
RL NUCLEIC ACIDS RES. 21:5408-5417(1993).
CC -!- CATALYTIC ACTIVITY: UROPORPHYRINOGEN III - COPROPORPHYRINOGEN +
CC 4 CO(2).
CC -!- PATHWAY: PORPHYRIN BIOSYNTHESIS.
DR EMBL; D12624; G216564; -
DR EMBL; U00006; G409791; -
DR EMBL; AE000473; G2367337; -.
DR PIR; JS0708; JS0708.
DR ECOGENE; EG11543; HEME.
DR PROSITE; PS00906; UROD_1; 1.
DR PROSITE; PS00907; UROD_2; 1.
KW LYASE; DECARBOXYLASE; PORPHYRIN BIOSYNTHESIS.
FT CONFLICT 72 78 AILFSDI -> RSSFRY (IN REF. 1).
FT CONFLICT 83 83 D -> I (IN REF. 1).
FT CONFLICT 89 103 LYFEAGEGPRTSPV -> SSILKPEKVRVLPQRI
FT (IN REF. 1)
FT CONFLICT 251 256 GGGQWL -> SATVA (IN REF. 1).
SQ SEQUENCE 354 AA; 59248 MW; A9638BF5 CRC32;

Query Match 66.7%; Score 44; DB 1; Length 354;
Best Local Similarity 57.1%; Pred. No. 2.04e+01;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 330 HQDVPE 336
||:|:|:|
QY 3 QOEVPDP 9

```

Search completed: Thu May 13 15:32:55 1999
Job time : 8 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Result No.	Score	Query Match %	Length	DB	ID	Description	Pred. No.
1	49	74.2	221	10	O65441	HYPOTHETICAL 24.8 KD P	6.28e+00
2	49	74.2	1406	4	OL5082	KIAA0377.	6.28e+00
3	49	74.2	1655	5	Q24754	MASTERMIND.	6.28e+00
4	48	72.7	1698	4	Q93021	PUTATIVE TUMOR SUPPRES	9.87e+00
5	47	71.2	147	2	O66475	SINGLE STRANDED DNA-BI	1.54e+01
6	47	71.2	216	11	O08904	BRX PROTEIN (FRAGMENT)	1.54e+01
7	47	71.2	1120	5	Q20778	SIMILAR TO TRIPLE HELI	1.54e+01
8	46	69.7	68	13	Q91773	HEAT SHOCK PROTEIN (HS	2.39e+01
9	46	69.7	264	5	Q24013	G5-LIKE ORF'S PROTEIN.	2.39e+01
10	46	69.7	360	10	O40645	OSB28.	2.39e+01
11	46	69.7	646	10	Q38870	CALMODULIN-DOMAIN PROT	2.39e+01
12	46	69.7	69	14	O71093	TERMINAL PROTEIN.	2.39e+01
13	46	69.7	812	5	Q18717	SIMILAR TO S.	2.39e+01
14	46	69.7	1320	14	O06359	150 KD PROTEIN.	2.39e+01
15	46	69.7	1828	14	O89249	209 KDA READTHROUGH PR	2.39e+01
16	46	69.7	7962	4	O14655	TITIN, SKELETAL MUSCLE	2.39e+01
17	45	68.2	370	5	Q02271	F52F12.2.	3.69e+01
18	45	68.2	405	6	P79260	S-ANTIGEN (FRAGMENT).	3.69e+01
19	45	68.2	411	5	Q26766	FLAGELLAR ANTIGEN (FRA	3.69e+01
20	45	68.2	974	5	P92165	ZK270.2C (FRAGMENT)	3.69e+01

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX NAGASE T., ISHIRAWA K., NAKAJIMA D., OHIRA M., SEKI N., MIYAJIMA N.,
RA TANAKA A., KOTANI H., NOMURA N., OHARA O.;
RL DNA RES. 4:141-150(1997);
DR EMBL; AB02375; D1021673;
SQ SEQUENCE 1406 AA; 156319 MW; 68F9BC7E CRC32;

Query Match 74.2%; Score 49; DB 4; Length 1406;
Best Local Similarity 55.6%; Pred. No. 6.28e+00;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 44 DEDEVPPPE 52
:|:|:|:|:
Qy 1 EEQEVPPD 9

RESULT 3
ID Q24754 PRELIMINARY; PRT: 1655 AA.
AC Q24754;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMREL. 01, LAST ANNOTATION UPDATE)
DE MASTERMIND.
OS DROSOPHILA VIRILIS. (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN [1]
RP SEQUENCE OF 260-762 FROM N.A.
RX MEDLINE; 91251140.
RA NEWFELD S.J., SMOLLER D.A., YEDVOBNICK B.;
RL J. MOL. EVOL. 32:415-420(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94111143.
RA NEWFELD S.J., SCHMID A.T., YEDVOBNICK B.;
RL J. MOL. EVOL. 37:483-495(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94365848.
RA NEWFELD S.J., TACHIDA H., YEDVOBNICK B.;
RL J. MOL. EVOL. 38:637-641(1994).
RN [4]
RP SEQUENCE FROM N.A.
RA NEWFELD S.J.;
RL SUBMITTED (MAY-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; M92914; G157834; -; Dvir\man.
DR FLYBASE; FBgn0013119; Dvir\man.
SQ SEQUENCE 1655 AA; 175048 MW; D521E17E CRC32;

Query Match 74.2%; Score 49; DB 5; Length 1655;
Best Local Similarity 55.6%; Pred. No. 6.28e+00;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 1084 QOQOQVPPN 1092
:|:|:|:|:
Qy 1 EEQEVPPD 9

RESULT 4
ID Q93021 PRELIMINARY; PRT: 698 AA.
AC Q93021;
DT 01-FEB-1997 (TREMREL. 02, CREATED)
DT 01-FEB-1997 (TREMREL. 02, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)
DE PUTATIVE TUMOR SUPPRESSOR.
GN LUCAL5.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-BRAIN;
RA SIMMLER M.C., HEARD E., ROUGEULLE C., CRUAUD C., WEISSENBAACH J.,

RN [1]
RP SEQUENCE FROM N.A.
RA BENILEY D., MAGGI L.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U73168; G1613900; -;
DR PFAM; PF00641; zf-RanBP.
SQ SEQUENCE 698 AA; 78066 MW; 36A3242B CRC32;

Query Match 72.7%; Score 48; DB 4; Length 698;
Best Local Similarity 60.0%; Pred. No. 9.87e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 94 DSEQEVPPGT 103
:|:|:|:|:
Qy 1 EEQEVPPDT 10

RESULT 5
ID O66475 PRELIMINARY; PRT: 147 AA.
AC O66475;
DT 01-AUG-1998 (TREMREL. 07, CREATED)
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE SINGLE STRANDED DNA-BINDING PROTEIN.
GN SSB.
OS AQUIFEX AEOLICUS.
OC EUBACTERIA; AQUIFACLES; AQUIFEX.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL NATURE 392:353-358(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AE000672; G2982816; -;
KW DNA-BINDING.
SQ SEQUENCE 147 AA; 17132 MW; BFA26F97 CRC32;

Query Match 71.2%; Score 47; DB 2; Length 147;
Best Local Similarity 75.0%; Pred. No. 1.54e+01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 115 EEEEVPP 122
|:|:|:|:|:
Qy 1 EEQEVPP 8

RESULT 6
ID O08904 PRELIMINARY; PRT: 216 AA.
AC O08904;
DT 01-JUL-1997 (TREMREL. 04, CREATED)
DT 01-JUL-1997 (TREMREL. 04, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE BRX PROTEIN (FRAGMENT).
GN BRX.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-BRAIN;
RA SIMMLER M.C., HEARD E., ROUGEULLE C., CRUAUD C., WEISSENBAACH J.,

RA AVNER P.;
RL MAMM. GENOME 0:0-0(0).
DR EMBL; Y11896; E311743; -
FT NON_TER 1
SQ SEQUENCE 216 AA; 24385 MW; 47AD381A CRC32;
Query Match 71.2%; Score 47; DB 11; Length 216;
Best Local Similarity 75.0%; Pred. No. 1.54e+01;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 76 EEEEEVPP 83
Qy 1 EEQEVPP 8
RESULT 7
ID Q20778 PRELIMINARY; PRT; 1120 AA.
AC Q20778;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE SIMILAR TO TRIPLE HELICAL REGION OF COLLAGENS.
GN F54B8.1.
OS CAENORHABDITIS ELSGANS.
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA BURTON J., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA WILSON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA VAUDIN M., VAUGHAN K., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RL NATURE 0:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U12966; G529221; -
SQ SEQUENCE 1120 AA; 119368 MW; E0352B8E CRC32;
Query Match 71.2%; Score 47; DB 5; Length 1120;
Best Local Similarity 50.0%; Pred. No. 1.54e+01;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
Db 127 EEEEEVPAEA 136
Qy 1 EEQEVPPDT 10
RESULT 8
ID Q91773 PRELIMINARY; PRT; 68 AA.
AC Q91773;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE HEAT SHOCK PROTEIN (HSP30) (FRAGMENT).
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AMPHIBIA; ANURA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84221917.

RA BIENZ M.;
RL PROC. NATL. ACAD. SCI. U.S.A. 81:3138-3142(1984).
DR EMBL; K02305; G214267; -
DR PFAM; PF00011; HSP20.
KW HEAT SHOCK.
FT NON_TER 1
SQ SEQUENCE 68 AA; 7405 MW; 4D1E326C CRC32;
Query Match 69.7%; Score 46; DB 13; Length 68;
Best Local Similarity 55.6%; Pred. No. 2.39e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 48 DAQEIPDDA 56
Qy 2 EQQEVPPDT 10
RESULT 9
ID Q24013 PRELIMINARY; PRT; 264 AA.
AC Q24013;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE G5-LIKE ORF'S PROTEIN.
OS DICTYOSTELIUM MUCOROIDES.
OC PLASMID DMP1; NUCLEAR.
OC EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DMUC2;
RX MEDLINE; 94302132.
RA KIYOSAWA H., HUGHES J.E., WELKER D.L.;
RL PLASMID 31:121-130(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-DMUC2;
RX MEDLINE; 98198836.
RA RIEBEN W.K. JR., GONZALES C.M., GONZALES S.T., PILKINGTON K.J.,
RA KIYOSAWA H., HUGHES J.E., WELKER D.L.;
RL GENETICS 148:1117-1125(1998).
DR EMBL; U00175; G2246436; -
KW PLASMID.
SQ SEQUENCE 264 AA; 30603 MW; 25AD28AD CRC32;
Query Match 69.7%; Score 46; DB 5; Length 264;
Best Local Similarity 62.5%; Pred. No. 2.39e+01;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 197 DEEEVPP 204
Qy 1 EQQEVPP 8
RESULT 10
ID Q40645 PRELIMINARY; PRT; 360 AA.
AC Q40645;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE OSB28.
OS ORYZA SATIVA (RICE).
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;
OC CYPERALES; GRAMINEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NIPPONBARE;
RX MEDLINE; 96417817.
RA NAKAGAWA H., OHMIYA K., HATTORI T.;
RL PLANT J. 9:217-227(1996).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER BZIP PROTEINS.
DR EMBL; U42208; G1147632; -

```
DR PROSITE; PS00036; BZIP_BASIC; 1.
DR PFAM; PF00170; BZIP.
KW DNA-BINDING; NUCLEAR PROTEIN.
SQ SEQUENCE 360 AA; 3893 MW; CBBACE0B CRC32;

Query Match
Best Local Similarity 69.7%; Score 46; DB 10; Length 360;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 298 DEAEVPP 305
   1 EEOQEVPP 8
   ||:|||||
   PRELIMINARY; PRT; 646 AA.

RESULT 11
ID Q38870 PRELIMINARY; PRT; 646 AA.
AC Q38870;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE CALMODULIN-DOMAIN PROTEIN KINASE CDPK ISOFORM 2.
GN CPK2.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
OC CAPPARALE; CRUCIFERAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA HRABAK E.M., DICKMANN L.J., SATTERLEE J.S., SUSSMAN M.R.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U31833; G1399271;
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PFAM; PF00036; ehand.
DR PFAM; PF00069; pkinase.
SQ SEQUENCE 646 AA; 72254 MW; 4F0F3FF3 CRC32;

Query Match
Best Local Similarity 69.7%; Score 46; DB 10; Length 646;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 105 ESKQEVPP 113
   1 EEOQEVPP 9
   ||:|||||
   PRELIMINARY; PRT; 649 AA.

RESULT 12
ID Q71093 PRELIMINARY; PRT; 649 AA.
AC Q71093;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE TERMINAL PROTEIN.
OS BOVINE ADENOVIRUS TYPE 3 (MASTADENOVIRUS BOS3).
OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; ADENOVIRIDAE; MASTADENOVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WBR-1;
RA BAXI M.K., REDDY P.S., ZAKHARTCHOUK A.N., IDAMAKANTI N., PYNE C.,
RA PYNE C., BABIUK L.A., TIKOO S.K.;
RA VIRUS GENES 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-WBR-1;
RA BAXI M.K., REDDY P.S., ZAKHARTCHOUK A.N., IDAMAKANTI N., PYNE C.,
RA BABIUK L.A., TIKOO S.K.;
RA VIRUS GENES 0:0-0(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-WBR-1;
RA REDDY P.S., IDAMAKANTI N., ZAKHARTCHOUK A.N., BAXI M.K., LEE J.B.,
RA PYNE C., BABIUK L.A., TIKOO S.K.;
RL J. VIROL. 72:1394-1402(1998).

[4]
RN SEQUENCE FROM N.A.
RP STRAIN-WBR-1;
RA REDDY P.S., IDAMAKANTI N., ZAKHARTCHOUK A.N., BAXI M.K., LEE J.B.,
RA PYNE C., BABIUK L.A., TIKOO S.K.;
RL SUBMITTED (OCT-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF030154; G2935217;
SQ SEQUENCE 649 AA; 75075 MW; 6BC9D410 CRC32;

Query Match
Best Local Similarity 69.7%; Score 46; DB 14; Length 649;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 391 EEEEDVPP 398
   1 EEOQEVPP 8
   ||:|||||
   PRELIMINARY; PRT; 812 AA.

RESULT 13
ID Q18717 PRELIMINARY; PRT; 812 AA.
AC Q18717;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE SIMILAR TO S.
DE CERVISIAE GENERAL NEGATIVE REGULATOR FO TRANSCRIPTION SUBUNIT 4.
GN C49H3.5.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACLOMOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSOUCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HANKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA KCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA WU X.;
RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA WATERSTON R.;
RL SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U42436; E214487;
SQ SEQUENCE 812 AA; 92341 MW; 132A227E CRC32;

Query Match
Best Local Similarity 69.7%; Score 46; DB 5; Length 812;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 518 EEOQIIPP 525
   1 EEOQEVPP 8
   ||:|||||
   PRELIMINARY; PRT; 1320 AA.

RESULT 14
ID Q06359 PRELIMINARY; PRT; 1320 AA.
AC Q06359;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-FEB-1997 (TREMBLREL. 02, LAST ANNOTATION UPDATE)
DE 150 KD PROTEIN.
OS SOIL-BORNE WHEAT MOSAIC VIRUS (SBWMV).
OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; NOT YET CLASSIFIED.
RN [1]
```

RP SEQUENCE FROM N.A.
RC STRAIN-US-NEBRASKA, 1981 WILD-TYPE;
RX MEDLINE: 93303914.
RA SHIRAKO Y., WILSON M.A.;
RL VIROLOGY 195:16-32(1993).
DR EMBL: L07937; G295508;
SQ SEQUENCE 1320 AA; 149976 MW; 58565281 CRC32;

Query Match 69.7%; Score 46; DB 14; Length 1320;
Best Local Similarity 44.4%; Pred. No. 2.39e+01;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 161 DEKEEIPPE 169
:|:|:|:
QY 1 EEQVEVPPD 9

RESULT 15
ID Q89249 PRELIMINARY; PRT; 1828 AA.
AC Q89249;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE 209 KDA READTHROUGH PROTEIN.
OS SOIL-BORNE WHEAT MOSAIC VIRUS.
OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; ROD-SHAPED SS-RNA VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-US-NEBRASKA, 1981 WILD-TYPE;
RX MEDLINE: 93303914.
RA SHIRAKO Y., WILSON M.A.;
RL VIROLOGY 195:16-32(1993).
DR EMBL: L07937; G295508;
SQ SEQUENCE 1828 AA; 209203 MW; ABA060DD CRC32;

Query Match 69.7%; Score 46; DB 14; Length 1828;
Best Local Similarity 44.4%; Pred. No. 2.39e+01;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 161 DEKEEIPPE 169
:|:|:|:
QY 1 EEQVEVPPD 9

Search completed: Thu May 13 15:33:44 1999
Job time : 30 secs.

M P S R C H

(TIM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu May 13 15:35:46 1999; Maspar time 2.54 Seconds
Tabular output not generated.
63.693 Million cell updates/sec

Title: >US-09-040-485-8
Description: (1-10) from US09040485.pep (1 of 2)
Perfect Score: 67
Sequence: 1 DGPTGEPQOE 10

Scoring table: PAM 150
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29

Statistics: Mean 14.719; Variance 44.118; scale 0.334

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	47	70.1	749 21	W14053	TIM01 mutant protein.	5.92e+01
2	47	70.1	1122 21	W14052	TIM protein.	5.92e+01
3	47	70.1	1389 21	W14051	TIM protein splice va	5.92e+01
4	47	70.1	3567 8	R44431	eryA region polypept	5.92e+01
5	45	67.2	1341 13	R71701	Collagen alpha 1 (I)	9.86e+01
6	45	67.2	1418 13	R71703	Collagen alpha 1 (II)	9.86e+01
7	45	67.2	1418 11	R59751	Type II collagen.	9.86e+01
8	44	65.7	243 10	W52601	ELA 243 amino acid pr	1.27e+02
9	44	65.7	745 20	R03560	Human adrenoleucodyst	1.27e+02
10	44	65.7	745 13	R76110	Human ALD	1.27e+02
11	43	64.2	943 28	W43039	A NADPH oxidase deriv	1.63e+02
12	43	64.2	3080 1	P93285	Sequence of clone HIV	1.63e+02
13	43	64.2	3210 1	P81771	Deduced sequence enco	1.63e+02
14	42	62.7	17 2	R06684	Human papilloma virus	2.09e+02
15	42	62.7	17 3	R14295	Seroreactive epitope	2.09e+02
16	42	62.7	19 5	R25268	HPV vaccine antigenic	2.09e+02
17	42	62.7	25 6	R31213	HPV-16 E7 peptide.	2.09e+02
18	42	62.7	98 29	W46886	Amino acid sequence o	2.09e+02

19	42	62.7	98 8	R42361	Human papillomavirus	2.09e+02
20	42	62.7	98 4	R22767	HPV E7 peptide.	2.09e+02
21	42	62.7	172 20	R97563	Human papilloma virus	2.09e+02
22	42	62.7	172 20	R97562	Human papilloma virus	2.09e+02
23	42	62.7	181 17	R94589	Human R33 platelet de	2.09e+02
24	42	62.7	181 17	R94588	Rat R33 platelet deri	2.09e+02
25	42	62.7	200 15	R76871	Talin C-terminal pept	2.09e+02
26	42	62.7	262 5	R27724	HPV 16 E7 protein fra	2.09e+02
27	42	62.7	266 20	R97561	Human papilloma virus	2.09e+02
28	42	62.7	324 24	W24971	Rat amelin-2 protein.	2.09e+02
29	42	62.7	407 24	W24970	Rat amelin-1 protein.	2.09e+02
30	42	62.7	433 25	W05389	Mouse SH3P7 protein.	2.09e+02
31	42	62.7	1442 14	R79480	Rat type II collagen.	2.09e+02
32	42	62.7	1931 5	R27649	Human calcium channel	2.09e+02
33	42	62.7	2237 14	R71006	Human neuronal calciu	2.09e+02
34	42	62.7	2237 6	R33550	Sequence of the alpha	2.09e+02
35	42	62.7	2339 14	R71005	Human neuronal calciu	2.09e+02
36	42	62.7	2339 6	R33549	Sequence of the alpha	2.09e+02
37	41	61.2	92 24	W14586	Streptococcus pneumon	2.68e+02
38	41	61.2	623 24	W12843	Pro-alpha1(III):(I) C	2.68e+02
39	41	61.2	626 24	W12842	Truncated pro-alpha1(2.68e+02
40	41	61.2	806 5	R27481	RP-III residual prote	2.68e+02
41	41	61.2	1070 14	R77083	M.vulgaris alpha-1,4-	2.68e+02
42	41	61.2	1070 13	R72712	Alpha-1,4-glucan lyas	2.68e+02
43	41	61.2	1078 13	R71704	Collagen alpha 1 (III	2.68e+02
44	41	61.2	1838 10	R53257	Human collagen (Type	2.68e+02
45	41	61.2	2353 19	R99393	Haemophilus adhesion	2.68e+02

ALIGNMENTS

RESULT 1
ID W14053 standard; Protein; 749 AA.
AC W14053;
DT 28-MAY-1997 (first entry)
DE TIM01 mutant protein.
KW TIM; timeless protein; nuclear translocation protein; circadian rhythm;
KW sleep-wake cycle; light sensitive; environmental cycle; nuclear protein;
KW Drosophila period gene; transcription factor; PAS domain; Depression;
KW narcolepsy; PER; jet lag; NTP; therapy.
OS Drosophila melanogaster.
FH Key Location/Qualifiers
FT region 351..380
FT /note="acidic region"
FT region 540..553
FT /note="basic region, possible nuclear localisation
FT signal"
PD WO9629406-A2.
PF 26-SEP-1996. U03830.
PR 20-MAR-1996; US-408518.
PR 16-MAY-1995; US-442214.
PR 02-NOV-1995; US-552354.
PR (UYPE-) UNIV PENNSYLVANIA.
PR (UYRQ) UNIV ROCKEFELLER.
PI Myers MP, Price JL, Sehgal A, Vossall LB, Young MW;
DR WPI; 96-443182/44.
DR N-PSDB; T60336.
PT Nuclear translocation protein which binds to protein involved in
PT circadian rhythms - used to develop prods. for diagnosis, prevention
PT or treatment of disorders associated with circadian rhythms, e.g.
PT jetlag and narcolepsy
PS Example 9; : 131pp; English.
CC This sequence represents the 01 mutant of the Drosophila melanogaster
CC "timeless" (TIM) protein. The TIM01 mutant is an arrhythmic mutation. TIM
CC is a nuclear translocation protein (NTP) of the invention. The NTPs of
CC the invention have specific binding activity to a protein involved in
CC circadian rhythms, and cyclic transcription patterns related to the
CC sleep-wake cycle. The NTP is also preferably light sensitive, and has a
CC stabilising effect on the circadian rhythm protein. The NTP also has the
CC ability to aid the process of circadian rhythm entrainment to
CC environmental cycles of light. TIM has specific binding activity for the
CC Drosophila period (PER) gene. PER is a nuclear protein which has homology

CC to the family of transcription factors containing the PAS domain, and is
 CC a protein involved in circadian rhythms, but the biological function of
 CC PER is unknown. The amount of PER fluctuates with a circadian rhythm.
 CC and the protein is also phosphorylated with a circadian rhythm. The NTPs
 CC of the invention, their fragments, agonists, mimics and antagonists are
 CC useful for preventing and/or treating disorders of a circadian rhythm
 CC such as depression, narcolepsy or jet lag. The products can also be used
 CC for detecting and/or measuring conditions so as to classify groups of
 CC individuals with the disorders.
 SQ Sequence 749 AA;

Query Match 70.1%; Score 47; DB 21; Length 749;
 Best Local Similarity 60.0%; Pred. No. 5.92e+01;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 510 dpgqgkphq 519
 ||| | | | | |
 QY 1 DGTGEPQOE 10

RESULT 2
 ID W14052 standard; Protein; 1122 AA.

AC W14052;
 DT 28-MAY-1997 (first entry)
 DE TIM protein.
 KW TIM; timeless protein; nuclear translocation protein; circadian rhythm;
 KW sleep-wake cycle; light sensitive; environmental cycle; nuclear protein;
 KW Drosophila period gene; transcription factor; PAS domain; depression;
 KW narcolepsy; PER; jet lag; NTP; therapy.
 OS Drosophila melanogaster.
 FH Key Location/Qualifiers
 FT region 351..380
 FT /note="acidic region"
 FT 540..553
 FT /note="basic region, possible nuclear localisation
 FT signal"

WO9629406-A2.
 26-SEP-1996.
 PD 20-MAR-1996; U03830.
 PF 20-MAR-1995; US-408518.
 PR 16-MAY-1995; US-442214.
 PR 02-NOV-1995; US-552354.
 PR (UYPE-) UNIV PENNSYLVANIA.
 PA (UYRQ) UNIV ROCKEFELLER.
 PI Myers MP, Price JL, Sehgal A, Voshall LB, Young MW;
 DR N-PSDB; T60333, T60335.
 DR WPI; 96-443182/44.
 PT Nuclear translocation protein which binds to protein involved in
 PT circadian rhythms - used to develop prods. for diagnosis, prevention
 PT or treatment of disorders associated with circadian rhythms, e.g.
 PT jetlag and narcolepsy
 PS Claim 3; Fig 11; 131pp; English.
 CC This sequence represents the Drosophila melanogaster "timeless" (TIM)
 CC protein. TIM is a nuclear translocation protein (NTP) of the invention.
 CC The NTPs of the invention have specific binding activity to a protein
 CC involved in circadian rhythms, and cyclic transcription patterns related
 CC to the sleep-wake cycle. The NTP is also preferably light sensitive, and
 CC has a stabilising effect on the circadian rhythm protein. The NTP also
 CC has the ability to aid the process of circadian rhythm entrainment to
 CC environmental cycles of light. TIM has specific binding activity for the
 CC Drosophila period (PER) gene. PER is a nuclear protein which has homology
 CC to the family of transcription factors containing the PAS domain, and is
 CC a protein involved in circadian rhythms, but the biological function of
 CC PER is unknown. The amount of PER fluctuates with a circadian rhythm.
 CC and the protein is also phosphorylated with a circadian rhythm. The NTPs
 CC of the invention, their fragments, agonists, mimics and antagonists are
 CC useful for preventing and/or treating disorders of a circadian rhythm
 CC such as depression, narcolepsy or jet lag. The products can also be used
 CC for detecting and/or measuring conditions so as to classify groups of
 CC individuals with the disorders.
 SQ Sequence 1122 AA;

Query Match 70.1%; Score 47; DB 21; Length 1122;

Best Local Similarity 60.0%; Pred. No. 5.92e+01;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Db 510 dpgqgkphq 519
 ||| | | | | |
 QY 1 DGTGEPQOE 10

RESULT 3
 ID W14051 standard; Protein; 1389 AA.

AC W14051;
 DT 28-MAY-1997 (first entry)
 DE TIM protein splice variant.
 KW TIM; timeless protein; nuclear translocation protein; circadian rhythm;
 KW sleep-wake cycle; light sensitive; environmental cycle; nuclear protein;
 KW Drosophila period gene; transcription factor; PAS domain; depression;
 KW narcolepsy; PER; jet lag; NTP; therapy.
 OS Drosophila melanogaster.
 FH Key Location/Qualifiers
 FT region 351..380
 FT /note="acidic region"
 FT 540..553
 FT /note="basic region, possible nuclear localisation
 FT signal"

WO9629406-A2.
 26-SEP-1996.
 PD 20-MAR-1996; U03830.
 PF 20-MAR-1995; US-408518.
 PR 16-MAY-1995; US-442214.
 PR 02-NOV-1995; US-552354.
 PR (UYPE-) UNIV PENNSYLVANIA.
 PA (UYRQ) UNIV ROCKEFELLER.
 PI Myers MP, Price JL, Sehgal A, Voshall LB, Young MW;
 DR N-PSDB; T60334.
 DR WPI; 96-443182/44.
 PT Nuclear translocation protein which binds to protein involved in
 PT circadian rhythms - used to develop prods. for diagnosis, prevention
 PT or treatment of disorders associated with circadian rhythms, e.g.
 PT jetlag and narcolepsy
 PS Claim 3; Fig 11; 131pp; English.
 CC This sequence represents the splice variant of the Drosophila
 CC melanogaster "timeless" (TIM) protein. TIM is a nuclear translocation
 CC protein (NTP) of the invention. The NTPs of the invention have specific
 CC binding activity to a protein involved in circadian rhythms, and cyclic
 CC transcription patterns related to the sleep-wake cycle. The NTP is also
 CC preferably light sensitive, and has a stabilising effect on the circadian
 CC rhythm protein. The NTP also has the ability to aid the process of
 CC circadian rhythm entrainment to environmental cycles of light. TIM has
 CC specific binding activity for the Drosophila period (PER) gene. PER is a
 CC nuclear protein which has homology to the family of transcription factors
 CC containing the PAS domain, and is a protein involved in circadian
 CC rhythms, but the biological function of PER is unknown. The amount of
 CC PER fluctuates with a circadian rhythm, and the protein is also
 CC phosphorylated with a circadian rhythm. The NTPs of the invention, their
 CC fragments, agonists, mimics and antagonists are useful for preventing
 CC and/or treating disorders of a circadian rhythm such as depression,
 CC narcolepsy or jet lag. The products can also be used for detecting and/or
 CC measuring conditions so as to classify groups of individuals with the
 CC disorders.
 SQ Sequence 1389 AA;

Query Match 70.1%; Score 47; DB 21; Length 1389;
 Best Local Similarity 60.0%; Pred. No. 5.92e+01;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 510 dpgqgkphq 519
 ||| | | | | |
 QY 1 DGTGEPQOE 10

RESULT 4
 ID R44431 standard; Protein; 3567 AA.
 AC R44431;

DT 22-DEC-1993 (first entry)
 DE eryA region polypeptide module #2.
 KW Saccaropolyspora erythraea; eryA; biosynthesis; polyketide; module;
 KW erythromycin; condensation; elongation; acyl chain growth;
 KW gene replacement.
 OS Saccaropolyspora erythraea.
 PN WO9313663-A.
 PD 22-JUL-1993.
 PF 17-JAN-1992; U00427.
 PR 17-JAN-1992; WO-U00427.
 PA (ABBO) ABBOTT LAB.
 PI Donadio S, Katz L, McAlpine JB;
 DR WPI: 93-242804/30.
 DR N-PSDB; Q46806.
 PT Biosynthesis of specific polyketide analogues esp. erythromycin
 PT cpds. - by introducing altered biosynthetic gene-contg. DNA into
 PT microorganisms
 PS Disclosure: Fig 2; 133pp; English.
 CC The sequences given in R4430-32 are encoded by the eryA fragment of
 CC the Saccaropolyspora erythraea genome. These polypeptides are
 CC involved in the biosynthesis of the polyketide segment of erythromycin.
 CC eryA is organised in modules and each module takes care of one
 CC condensation step. The precise succession of elongation steps is
 CC dictated by the genetic order of the modules. The DNA encoding
 CC these polypeptides may be specifically altered such that novel
 CC polyketide molecules of desired structure are produced. Three types
 CC of alteration may be produced; those inactivating a single function in
 CC a module which does not arrest acyl chain growth; those inactivating a
 CC single function in a module which does affect chain growth; and those
 CC affecting an entire module. The mutations may be introduced by gene
 CC replacement.
 CC Sequence 3567 AA;
 SQ

Query Match 70.1%; Score 47; DB 8; Length 3567;
 Best Local Similarity 87.3%; Pred. No. 5.92e+01;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1376 gptgepaq 1383
 QY 2 GPTGEPOQ 9

RESULT 5
 ID R71701 standard; protein; 1341 AA.
 AC R71701;
 DT 17-OCT-1995 (first entry)
 DE Collagen alpha 1 (I) chain precursor.
 KW Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring;
 KW disorder; osteoporosis; metastatic progression; Paget's disease;
 KW hyperthyroidism; bone; resorption; rheumatoid arthritis;
 KW osteoarthritis; vasculitis syndrome.
 OS Homo sapiens.
 FT Key Location/Qualifiers
 FT misc_difference 2028
 FT /note= "Unidentified amino acid."
 PN WO9508115-A.
 PD 23-MAR-1995.
 PF 19-SEP-1994; DK0348.
 PR 17-SEP-1993; DK-001040.
 PA (OSTE-) OSTEOMETER AS.
 PI Bonde M, Qvist P;
 DR WPI: 95-131456/17.
 PT Assaying collagen fragments in body fluid by immunoassay - using
 PT antibodies raised against synthetic peptide(s) contg. potential
 PT crosslinking sites, to diagnose and monitor disorders of collagen
 PT metabolism, e.g. osteoporosis.
 PS Disclosure (Appendix A); Page 49; 87pp; English.
 CC Determination of collagen fragments in body fluids can be achieved
 CC by immunoassay using antibodies directed against synthetic peptides
 CC derived from collagen which contain sites of potential crosslinking.
 CC The method is used to diagnose and monitor treatment of disorders of
 CC collagen metabolism (degradation of type I collagen may indicate
 CC osteoporosis, metastatic progression, Paget's disease,

CC hyperthyroidism or other conditions involving excessive bone
 CC resorption; degradation of type II collagen may indicate rheumatoid
 CC arthritis or osteoarthritis; and of type III collagen, vaculitis
 CC syndrome). The method can also be used to assess the toxicity of a
 CC compound and to test drugs for their effect on collagen metabolism.
 SQ Sequence 1341 AA;

Query Match 67.2%; Score 45; DB 13; Length 1341;
 Best Local Similarity 44.4%; Pred. No. 9.86e+01;
 Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 202 zgpgpgpgz 210
 QY 1 DGPTGEPOQ 9

RESULT 6
 ID R71703 standard; protein; 1418 AA.
 AC R71703;
 DT 17-OCT-1995 (first entry)
 DE Collagen alpha 1 (II) chain precursor.
 KW Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring;
 KW disorder; osteoporosis; metastatic progression; Paget's disease;
 KW hyperthyroidism; bone; resorption; rheumatoid arthritis;
 KW osteoarthritis; vasculitis syndrome.
 OS Homo sapiens.
 PN WO9508115-A.
 PD 23-MAR-1995.
 PF 19-SEP-1994; DK0348.
 PR 17-SEP-1993; DK-001040.
 PA (OSTE-) OSTEOMETER AS.
 PI Bonde M, Qvist P;
 DR WPI: 95-131456/17.
 PT Assaying collagen fragments in body fluid by immunoassay - using
 PT antibodies raised against synthetic peptide(s) contg. potential
 PT crosslinking sites, to diagnose and monitor disorders of collagen
 PT metabolism, e.g. osteoporosis.
 PS Disclosure (Appendix A); Page 53; 87pp; English.
 CC Determination of collagen fragments in body fluids can be achieved
 CC by immunoassay using antibodies directed against synthetic peptides
 CC derived from collagen which contain sites of potential crosslinking.
 CC The method is used to diagnose and monitor treatment of disorders of
 CC collagen metabolism (degradation of type I collagen may indicate
 CC osteoporosis, metastatic progression, Paget's disease,
 CC hyperthyroidism or other conditions involving excessive bone
 CC resorption; degradation of type II collagen may indicate rheumatoid
 CC arthritis or osteoarthritis; and of type III collagen, vaculitis
 CC syndrome). The method can also be used to assess the toxicity of a
 CC compound and to test drugs for their effect on collagen metabolism.
 SQ Sequence 1418 AA;

Query Match 67.2%; Score 45; DB 13; Length 1418;
 Best Local Similarity 66.7%; Pred. No. 9.86e+01;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 930 gpgpgpgq 938
 QY 2 GPTGEPOQ 10

RESULT 7
 ID R59751 standard; protein; 1418 AA.
 AC R59751;
 DT 14-FEB-1995 (first entry)
 DE Type II collagen.
 KW Collagen; triple helix; articular cartilage; collagenase;
 KW degradation; monoclonal antibody; epitope; matrix;
 KW metalloproteinase.
 OS Homo sapiens.
 FT Key Location/Qualifiers
 FT peptide 1..24
 FT /label= Signal peptide.
 PN WO9414070-A.

PD 23-JUN-1994.
 PF 06-DEC-1993; CA0522.
 PR 04-DEC-1992; US-984123.
 PA (SHRI-) SHRINERS HOSPITALS FOR CRIPPLED CHILDREN.
 PI Hollander AB, Poole AR;
 DR WPI: 94-234222/28.
 PT Determin. of cartilage degradation - using a monoclonal antibody
 to measure the amt. of unwound collagen or fragments in samples
 PS Disclosure; Figure 1; 119pp; English.
 CC Type II collagen constitutes the bulk of the fibrillar backbone of
 cartilage matrix. It is composed of a tightly wound triple helix
 which can only be cleaved by the metalloproteinase collagenase to
 produce 3/4 and 1/4 length alpha chain fragments. The destruction of
 articular cartilage is due, in part, to the degradation of collagen.
 CC Incapable of maintaining its helical structure at physiological
 temperatures, collagenase-cleaved collagens unwind and become
 susceptible to further degradation by other proteinases. By
 CC producing monoclonal antibodies directed against epitopes which are
 only revealed when collagen is unwound, the antibodies provide a
 CC means of determining the degradation of cartilage in a biological
 CC sample. The antibodies do not bind to native helical collagen.
 CC Epitopes used in the production of such antibodies are described in
 CC R59749, R59750 and R67742.
 SQ Sequence 1418 AA;

Query Match 67.2%; Score 45; DB 11; Length 1418;
 Best Local Similarity 66.7%; Pred. No. 9.86e+01;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 930 gspgpgqgq 938
 QY 2 GPTGEPOQE 10
 :|||:|

RESULT 8
 ID R52601 standard; Protein; 243 AA.
 AC R52601;
 DT 05-DEC-1994 (first entry)
 DE EIA 243 amino acid protein.
 KW Alternative splicing; EIA; adenovirus; early gene; oncogenes;
 transcription regulation protein; oncogenic transformation; rodent;
 KW phenotype; human; tumour cell; contact inhibition; differentiation;
 adenovirus 5; anchorage-independent growth; tumorigenic potential;
 KW reorganisation; flat morphology; tumour suppressor.
 OS Adenovirus.
 PN W09409160-A.
 PD 28-APR-1994.
 PF 13-OCT-1993; U09774.
 PR 13-OCT-1992; US-960112.
 PA (LJOL-) LA JOLLA CANCER RES FOUND.
 PI Frisch SW;
 DR WPI: 94-151344/18.
 DR N-PSDB; 062526.
 PT Restoring contact inhibition to hyperproliferative cells - by
 introducing nucleic acid encoding EIA peptide, also promoting
 PT differentiation, used for treating malignancies
 PS Disclosure; Page 20-22; 41pp; English.
 CC This sequence is encoded by the product of alternative splicing of the
 CC EIA RNA such that the information contained within this sequence is a
 CC subset of the larger EIA protein, see also R52602. EIA is an adeno-
 CC virus early gene and produces two products. The 243 and 289 residue
 CC proteins are both transcriptional regulation proteins which facilitate
 CC the oncogenic transformation of certain rodent cells by other
 CC oncogenes. The adenovirus EIA gene unexpectedly influences the
 CC phenotype of human tumour cells so as to restore their contact
 CC inhibitory properties and promote differentiation. Stable-expression
 CC of the adenovirus 5 EIA gene reduces anchorage-independent growth and
 CC tumorigenic potential, promotes reorganisation, induces flat
 CC morphology, and restores contact inhibition in human tumour cell
 CC lines. Therefore EIA acts as a tumour suppressor gene in this human
 CC context.
 SQ Sequence 243 AA;

Query Match 65.7%; Score 44; DB 10; Length 243;
 Best Local Similarity 50.0%; Pred. No. 1.27e+02;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 138 egpvsepepe 147
 QY 1 DGPTGEPOQE 10
 :||:|:|

RESULT 9
 ID W03560 standard; Protein; 745 AA.
 AC W03560;
 DT 26-FEB-1997 (first entry)
 DE Human adrenoleucodystrophy wild-type protein ALDP.
 KW Adrenoleucodystrophy; gene therapy; retroviral vector M48;
 KW adrenomyelopathy; membrane protein; long chain fatty acid oxidation.
 OS Homo sapiens.
 PN W09621733-A2.
 PD 18-JUL-1996.
 PF 12-JAN-1996; F00059.
 PR 13-JAN-1995; FR-000376.
 PA (INRM) INST NAT SANTE & RECH MEDICALE.
 PI Aubourg P, Mandel JL, Mosser J, Sarde CO;
 DR WPI: 96-342286/34.
 DR N-PSDB; T39335.
 PT Recombinant viral vector contg. DNA for correcting
 PT adrenoleuco-dystrophy - and immunologically tolerable cells contg.
 PT this vector, useful in gene therapy
 PS Example; Fig 7; 36pp; French.
 CC A 2.43 kb SpeI-EcoRI fragment of human wild-type ALD
 CC (adrenoleucodystrophy) cDNA was inserted into retroviral vector M48
 CC so that it was under control of the mouse phosphoglycerate kinase
 CC (PGK) promoter. The resulting vector was co-transfected with vector
 CC PGK-neomycin into amphotropic packaging cell line psiCRIP. Neomycin
 CC (G418)-resistant transfectant cells which were also positive with
 CC anti-ALD antibodies were incubated with skin fibroblasts of an ALD
 CC patient. After two rounds of infection, about 70% of cells were
 CC expressing normal ALD protein (ALDP) and integration of M48-ALD was
 CC confirmed by Southern blotting. The present sequence is that of
 CC ALDP.
 SQ Sequence 745 AA;

Query Match 65.7%; Score 44; DB 20; Length 745;
 Best Local Similarity 60.0%; Pred. No. 1.27e+02;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 47 qapageptqe 56
 QY 1 DGPTGEPOQE 10
 :|||:|

RESULT 10
 ID R76110 standard; Protein; 745 AA.
 AC R76110; 1995 (first entry)
 DT 16-NOV-1995 (first entry)
 DE Human ALD.
 KW ALD gene; adrenoleukodystrophy; adrenomyeloneuropathy; AMN;
 KW diagnosis; gene therapy.
 OS Homo sapiens.
 PN CA2108606-A.
 PD 16-APR-1995.
 PF 15-OCT-1993; 108606.
 PR 15-OCT-1993; CA-108606.
 PA (INRM) INST NAT SANTE & RECH MEDICALE.
 PI Aubourg P, Mandel J, Mosser J, Sarde C;
 DR WPI: 95-215721/29.
 DR N-PSDB; Q94048.
 PT New nucleic acid responsible for adreno-leuco-dystrophy - related
 PT probes, proteins and antibodies, useful for diagnosis and treatment
 PS Claim 16; Fig.2A-B; 39pp; English.
 CC Probes corresp. to breakpoints in the red pigment gene of an AMN
 CC patient were used to isolate clones from an Xq28 cosmid library.
 CC Following hybridizations and nested PCR, a HeLa cell cDNA library

CC was screened, and a complete ALD cDNA fragment (Q94048) encoding the protein given in R94048 was obt'd. Transformation of hematopoietic cells with ALD-encoding sequences in vivo will allow therapy of ALD or AMN.

CC Sequence 745 AA;

Query Match 65.78; Score 44; DB 13; Length 745;
Best Local Similarity 60.0%; Pred. No. 1.27e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 47 gapageptqe 56
:::|||||

QY 1 DGPTGEPQOE 10

RESULT 11

ID W43039 standard; Protein; 943 AA.

AC W43039;

DT 12-MAY-1998 (first entry)

DE A NADPH oxidase derived from Arabidopsis thaliana.

KW NADPH oxidase; development; agricultural chemical; growth control;

KW transformation; plant; evaluation.

OS Arabidopsis thaliana.

FH Key Location/Qualifiers

FT Misc_difference 161 /note= "not specified"

FT Misc_difference 165 /note= "not specified"

FT

PN J10033176-A.

PD 10-FEB-1998.

PF 23-JUL-1996; 193220.

PR 23-JUL-1996; JP-193220.

PA (SUMO) SUMITOMO CHEM CO LTD.

DR WPI: 98-172095/16.

DR N-PSDB; V03045.

PT NADPH oxidase derived from plants - useful for control of growth of transformed plants and for evaluation of agricultural chemicals

PS Claim 1; Pages 7-9; 11pp; Japanese.

CC The presents sequence represents a NADPH oxidase. The cDNA encoding this protein sequence was isolated from a cDNA library of Arabidopsis thaliana. The NADPH DNA and protein can be used in the development of agricultural chemicals. They allow for the control of growth of transformed plants for smooth evaluation of agricultural chemicals.

CC Sequence 943 AA;

Query Match 64.28; Score 43; DB 28; Length 943;
Best Local Similarity 60.0%; Pred. No. 1.63e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 732 dgpygapagd 741
|||||

QY 1 DGPTGEPQOE 10

RESULT 12

ID P93285 standard; Protein; 3080 AA.

AC P93285;

DT 06-APR-1990 (first entry)

DE Sequence of clone HIV-2 SBL/ISY.

KW HIV-2; proviral clone HIV-2 SBL/ISY.

OS Human immunodeficiency virus 2.

PN US7331212-A.

PD 29-AUG-1989.

PF 31-MAR-1989; 331212.

PR 31-MAR-1989; US-331212.

PA (USSH) US Dept. Health and Human Services.

PI Franchini G, Wong-Staal F, Gallo R;

DR WPI: 89-339698/46.

DR N-PSDB; N92119.

PT Complete human immunodeficiency type 2 proviral clone - used to generate animal model for function studies of HIV genes in vivo.

PS Disclosure; Fig. 5; 43pp; English.

CC The protein is encoded by the third reading frame of HIV-2 SBL/ISY, a

CC proviral clone of HIV-2.

SQ Sequence 3080 AA;

Query Match 64.28; Score 43; DB 1; Length 3080;
Best Local Similarity 75.0%; Pred. No. 1.63e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 2799 phgepqqq 2806
|||||

QY 3 PTGEPQOE 10

RESULT 13

ID P81771 standard; Protein; 3210 AA.

AC P81771;

DT 07-NOV-1990 (first entry)

DE Deduced sequence encoded by bottom reading frame of cDNA clone HIV-2 SBL/ISY of HIV related retrovirus strain

KW HIV vaccine; HIV strain SBL-6669-85.

OS Human immunodeficiency virus.

PN W08080449-A.

PD 03-NOV-1988.

PF 28-APR-1988; SE0218.

PR 28-APR-1987; SE-001765.

PA (SBLs-) SBL Statens Bekteri.

PI Albert J, Biberfeld G, Fenyo EM, Norrby E;

DR WPI: 88-322769/45.

DR N-PSDB; n80890.

PT HIV related human retro-virus strain - used for obtaining antigens for assays and vaccines and for

PT Claim. of antibodies for assays

PS Claim 9; Fig 4; 28pp; English.

CC Synthetically produced proteins and peptides, characterised in that the AA sequence is derived from the primary nucleotide sequence of HIV-2 SBL/ISY or a part thereof, or a degenerate thereof are claimed.

CC HIV-2 SBL/ISY represents the complete genome of the virus SBL-6669 (-SBL-6669-85). The proviral DNA was obt'd. from a genomic library constructed from DNA of HUT-78 cells infected with SBL-6669-85 using the lambda-phage vector EMBL-3. SBL-6669-85 was isolated from lymphocytes of a West African woman. Protection is requested for the entire genome disclosed in n80890 and for parts thereof, and corresp. to various genes such as the gag gene (corresp. to nucleotides 547 to 2106), the pol gene (nucleotides 1827-4931) and the env gene (nucleotides 6144 to 8682), the corresp. AA sequences and parts thereof and various products derived therefrom, or use thereof, such as clones prep'd. by recombinant vector method, HIV test devices and methods. X corresponds to the translation of a stop codon.

CC Sequence 3210 AA;

Query Match 64.28; Score 43; DB 1; Length 3210;
Best Local Similarity 75.0%; Pred. No. 1.63e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 2906 phgepqqq 2913
|||||

QY 3 PTGEPQOE 10

RESULT 14

ID R06684 standard; Protein; 17 AA.

AC R06684;

DT 11-JAN-1991 (first entry)

DE Human papilloma virus 16 E7-protein DNA sequence (V).

KW Immunogenic region; Human Papilloma Virus; Hpv16 E7; diagnosis;

KW antibodies; vaccines.

OS Human Papilloma Virus 16 E7.

PN EP-386734-A.

PD 12-SEP-1990.

PF 07-MAR-1990; 104353.

PR 10-MAR-1989; DE-907721.

PA (BEHW) BEHRINGERWERKE AG.

PI Bartsch D, Gissmann L, Muller M;

DR WPI: 90-276785/37.

Search completed: Thu May 13 15:36:03 1999
Job time : 17 secs.

 WIREH (TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Thu May 13 15:40:37 1999; MasPar time 2.71 Seconds
 Tabular output not generated. 82.989 Million cell updates/sec

Title: >US-09-040-485-8
 Description: (3-8) from US09040485.pep (2 of 2)
 Perfect Score: 41
 Sequence: 1 PTGEPQ 6

Scoring table: PAM 150
 GAP 15

Searched: 116738 seqs, 37463448 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: pir58
 1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 17.462; Variance 17.910; scale 0.975

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	41	100.0	115	2	A23925 proline-rich phosphoprotein	1.72e+00
2	41	100.0	757	2	I38423 aspartyl beta-hydroxylase	1.72e+00
3	38	92.7	232	1	MMB213 25.5k membrane protein	9.96e+00
4	38	92.7	484	1	SYR2ET glutamate--tRNA ligas	9.96e+00
5	37	90.2	325	1	B48561 inner capsid protein	1.75e+01
6	37	90.2	325	1	A48561 inner capsid protein	1.75e+01
7	37	90.2	328	1	JQ1875 inner capsid protein	1.75e+01
8	37	90.2	328	1	VPXRC3 minor inner core prot	1.75e+01
9	37	90.2	369	2	A36909 C4-dicarboxylase-bind	1.75e+01
10	37	90.2	370	2	A36909 glutamyl-tRNA synthet	1.75e+01
11	37	90.2	468	2	S21172 glutamate--tRNA ligas	1.75e+01
12	37	90.2	494	2	S76516 integral membrane pro	1.75e+01
13	37	90.2	505	2	JC5208 glutamate--tRNA ligas	1.75e+01
14	37	90.2	506	2	F71513 probable glutamyl-tRN	1.75e+01
15	36	87.8	142	1	SKPSXT secretion protein xcp	3.05e+01
16	36	87.8	612	2	G69649 GTP-binding protein 1	3.05e+01
17	36	87.8	770	2	S60876 cellobiose oxidase (E	3.05e+01
18	35	85.4	231	2	S76204 hypothetical protein	5.26e+01
19	35	85.4	235	2	S23249 transcription factor	5.26e+01
20	35	85.4	231	1	PIHUPF salivary proline-rich	5.26e+01
21	35	85.4	303	2	C70808 hypothetical protein	5.26e+01
22	35	85.4	324	2	SL1497 CAMP-binding protein	5.26e+01
23	35	85.4	333	2	SL1484 CABP1 protein - slime	5.26e+01

24	35	85.4	337	2	A37760 UDPglucose 4-epimeras	5.26e+01
25	35	85.4	392	1	PIHUB6 salivary proline-rich	5.26e+01
26	35	85.4	441	2	S74800 dihydroorotase (EC 3.	5.26e+01
27	35	85.4	512	2	SL9354 cell fusion protein F	5.26e+01
28	35	85.4	535	2	P70782 probable oxidoreducta	5.26e+01
29	35	85.4	545	2	A39193 cytochrome c3 precu	5.26e+01
30	35	85.4	562	2	A41707 oligo-1,6-glucosidase	5.26e+01
31	35	85.4	600	2	H70448 G-protein Lepa - Aqu	5.26e+01
32	35	85.4	622	2	I48733 protein kinase rck (E	5.26e+01
33	35	85.4	622	2	A34711 kinase-related protei	5.26e+01
34	35	85.4	666	2	B70803 hypothetical protein	5.26e+01
35	35	85.4	2336	2	A45386 omega-conotoxin-sensi	5.26e+01
36	35	85.4	3712	2	S18253 laminin alpha-1 chain	5.26e+01
37	34	82.9	96	3	JC5931 high mobility group I	8.97e+01
38	34	82.9	96	3	JC5933 high mobility group I	8.97e+01
39	34	82.9	105	3	JC5932 high mobility group I	8.97e+01
40	34	82.9	436	2	S69821 NDP-sugar dehydrogena	8.97e+01
41	34	82.9	597	2	G71373 probable methyl-accept	8.97e+01
42	34	82.9	899	1	A43969 core protein VP3 - ep	8.97e+01
43	34	82.9	901	2	S07419 core protein P3 - blu	8.97e+01
44	34	82.9	1838	1	CGHUIV collagen alpha 1(V) c	8.97e+01
45	34	82.9	1843	2	SL8803 collagen alpha 1(V) c	8.97e+01

ALIGNMENTS

RESULT 1
 ENTRY proline-rich phosphoprotein - crab-eating macaque
 TITLE #formal_name Macaca fascicularis #common_name crab-eating macaque
 ORGANISM 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 25-Oct-1996

ACCESSIONS A23925
 REFERENCE A23925
 #authors Oppenheim, F.G.; Offner, G.D.; Troxler, R.F.
 #journal J Biol Chem 263:10611-10628
 #title Amino acid sequence of a proline-rich phosphoglycoprotein from parotid secretion of the subhuman primate Macaca fascicularis.

#cross-references MUID:85289254
 #accession A23925
 #molecule_type protein
 #residues 1-115 #label OPP
 CLASSIFICATION #superfamily proline-rich protein
 KEYWORDS phosphoprotein
 SUMMARY #length 115 #molecular-weight 12255 #checksum 2737

Query Match 100.0%; Score 41; DB 2; Length 115;
 Best Local Similarity 100.0%; Pred. No. 1.72e+00;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 68 PTGEPQ 73

QY 3 PTGEPQ 8

RESULT 2

ENTRY aspartyl beta-hydroxylase - human
 TITLE #formal_name Homo sapiens #common_name man
 ORGANISM 29-May-1998 #sequence_revision 29-May-1998 #text_change 10-Jul-1998
 DATE I38423
 ACCESSIONS I38423
 REFERENCE Koriath, F.; Gieffers, C.; Frey, J.
 #authors Gene (1994) 150:395-399
 #journal Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase.
 #title aspartyl beta-hydroxylase.
 #cross-references MUID:95121937
 #accession I38423
 #status Preliminary; translated from GB/EMBL/DBJ
 #molecule_type mRNA

```

##residues 1-757 ##label RES
##cross-references EMBL:U03109; NID:g458031; PID:g458032
CLASSIFICATION #superfamily peptide-aspartate beta-dioxygenase;
tetrairicopeptide repeat homology
FEATURE
54-75 #domain transmembrane #status predicted #label TRM
SUMMARY #length 757 #molecular-weight 85498 #checksum 2143

Query Match 100.0%; Score 41; DB 2; Length 757;
Best Local Similarity 100.0%; Pred. No. 1.72e+00;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 175 PTGEPQ 180
|||||
Qy 3 PTGEPQ 8

RESULT 3
ENTRY MMBEI3 #type complete
TITLE 25.5K membrane protein - ictalurid herpesvirus 1 (strain
aurn 1)
ORGANISM #formal_name ictalurid herpesvirus 1
#note host Ictalurus punctatus (channel catfish)
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change
05-Sep-1997
ACCESSIONS I36786
REFERENCE A36804
#authors Davison, A.J.
#submission submitted to GenBank, January 1992
#description Channel catfish virus: a new type of herpesvirus.
#accession I36786
#molecule_type DNA
##residues 1-232 ##label DAV
##cross-references GB:M75136; NID:g331209; PID:g331218
REFERENCE A39447
#authors Davison, A.J.
#journal Virology (1992) 186:9-14
#title Channel catfish virus: a new type of herpesvirus.
#cross-references MUID:92087490
#contents annotation
#note neither amino acid nor nucleotide sequence is given
GENETICS
#gene 8
CLASSIFICATION #superfamily ictalurid herpesvirus 25.5K membrane protein
KEYWORDS transmembrane protein
FEATURE
166-182 #domain transmembrane #status predicted #label TM1\
195-211 #domain transmembrane #status predicted #label TM2
SUMMARY #length 232 #molecular-weight 25466 #checksum 8650

Query Match 92.7%; Score 38; DB 1; Length 232;
Best Local Similarity 83.3%; Pred. No. 9.96e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 90 PTGEPH 95
|||||
Qy 3 PTGEPH 8

RESULT 4
ENTRY SYRZET #type complete
TITLE glutamate--tRNA ligase (EC 6.1.1.17) - Rhizobium meliloti
ALTERNATE_NAMES glutamyl-tRNA synthetase
ORGANISM #formal_name Rhizobium meliloti
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
05-Sep-1997
ACCESSIONS A32888
REFERENCE A32888
#authors Laberge, S.; Gagnon, Y.; Bordeleau, L.M.; Lapointe, J.
#journal J. Bacteriol. (1989) 171:3926-3932
#title Cloning and sequencing of the gltx gene, encoding the
glutamyl-tRNA synthetase of Rhizobium meliloti A2.
#cross-references MUID:89291743

#accession A32888
#molecule_type DNA
#residues 1-325 ##label LAB
#cross-references GB:M27221; NID:g341649; PID:g717082
##experimental_source strain A2
CLASSIFICATION #superfamily glutamate--tRNA ligase
KEYWORDS aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis
FEATURE
256 #binding_site ATP (Lys) #status predicted
SUMMARY #length 484 #molecular-weight 54228 #checksum 7829

Query Match 92.7%; Score 38; DB 1; Length 484;
Best Local Similarity 83.3%; Pred. No. 9.96e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 14 PTGEPH 19
|||||
Qy 3 PTGEPH 8

RESULT 5
ENTRY B48561 #type complete
TITLE inner capsid protein VP6 - bluetongue virus (serotype 10,
strain USA)
ORGANISM #formal_name bluetongue virus
DATE 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change
20-Mar-1998
ACCESSIONS B48561
REFERENCE A48561
#authors Hwang, G.Y.; Chiou, J.F.; Yang, Y.Y.; Li, J.K.
#journal Virus Res. (1992) 24:315-323
#title Comparative sequence analyses of the cognate structural
protein VP6 genes of five US bluetongue viruses.
#cross-references MUID:93033709
#accession B48561
#molecule_type genomic RNA
#residues 1-325 ##label HWA
#cross-references GB:L08669; NID:g210842; PID:g210843
#note sequence extracted from NCBI backbone (NCBIP:115445)
GENETICS
#map_position segment 9
CLASSIFICATION #superfamily bluetongue virus VP6 protein
KEYWORDS capsid protein
SUMMARY #length 325 #molecular-weight 35163 #checksum 4288

Query Match 90.2%; Score 37; DB 1; Length 325;
Best Local Similarity 66.7%; Pred. No. 1.75e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 285 PTGDPH 290
|||||
Qy 3 PTGDPH 8

RESULT 6
ENTRY A48561 #type complete
TITLE inner capsid protein VP6 - bluetongue virus (serotype 11,
strain USA)
ORGANISM #formal_name bluetongue virus
DATE 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change
17-Feb-1994
ACCESSIONS A48561
REFERENCE A48561
#authors Hwang, G.Y.; Chiou, J.F.; Yang, Y.Y.; Li, J.K.
#journal Virus Res. (1992) 24:315-323
#title Comparative sequence analyses of the cognate structural
protein VP6 genes of five US bluetongue viruses.
#cross-references MUID:93033709
#accession A48561
#molecule_type genomic RNA
#residues 1-325 ##label HWA
#note sequence extracted from NCBI backbone (NCBIP:115444)

```

GENETICS

#map_position segment 9
 CLASSIFICATION #superfamily bluetongue virus VP6 protein
 KEYWORDS capsid protein
 SUMMARY #length 325 #molecular-weight 35354 #checksum 2927

Query Match 90.2%; Score 37; DB 1; Length 325;
 Best Local Similarity 66.7%; Pred. No. 1.75e+01;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 285 PTGDPH 290
 |||:
 Qy 3 PTGEPQ 8

RESULT 7

ENTRY JQ1875 #type complete
 TITLE inner capsid protein VP6 - bluetongue virus (serotype 1, strain South Africa)
 ALTERNATE_NAMES VP6 protein
 ORGANISM #formal_name bluetongue virus
 DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Sep-1997

ACCESSIONS

REFERENCE JQ1875
 #authors Wade-Evans, A.M.; Mertens, P.P.C.; Belsham, G.J.

#journal J. Gen. Virol. (1992) 73:3023-3026
 #title Sequence of genome segment 9 of bluetongue virus (serotype 1, South Africa) and expression analysis demonstrating that different forms of VP6 are derived from initiation of protein synthesis at two distinct sites.

#accession JQ1875

#molecule_type genomic RNA
 #residues 1-328 #label WAD
 #cross-references GB:D10905; NID:G221080; PID:d1002188; PID:g221081

GENETICS

#map_position segment 9
 CLASSIFICATION #superfamily bluetongue virus VP6 protein
 KEYWORDS capsid protein
 SUMMARY #length 328 #molecular-weight 35928 #checksum 4514

Query Match 90.2%; Score 37; DB 1; Length 328;
 Best Local Similarity 66.7%; Pred. No. 1.75e+01;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294
 |||:
 Qy 3 PTGEPQ 8

RESULT 8

ENTRY VPXRC3 #type complete
 TITLE minor inner core protein VP6 - bluetongue virus (serotype 10, American isolate)
 ALTERNATE_NAMES inner capsid protein VP6; VP6 protein
 ORGANISM #formal_name bluetongue virus
 DATE 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 10-Nov-1995

ACCESSIONS B32400; S10542
 REFERENCE A32400

#authors Fukusho, A.; Yu, Y.; Yamaguchi, S.; Roy, P.

#journal J. Gen. Virol. (1989) 70:1677-1689
 #title Completion of the sequence of bluetongue virus serotype 10 by the characterization of a structural protein, VP6, and a non-structural protein, NS2.

#cross-references MUID:89293076

#accession B32400
 #molecule_type genomic RNA
 #residues 1-328 #label FUK

REFERENCE

S10534
 #authors Roy, P.; Marshall, J. J. A.; French, T. J.
 #journal Curr. Top. Microbiol. Immunol. (1990) 162:43-87
 #title Structure of the bluetongue virus genome and its encoded

proteins.

#cross-references MUID:90345726
 #accession S10542
 #status preliminary
 #molecule_type genomic RNA
 #residues 1-328 #label CUR

GENETICS

#map_position segment 9
 CLASSIFICATION #superfamily bluetongue virus VP6 protein
 KEYWORDS capsid protein
 SUMMARY #length 328 #molecular-weight 35749 #checksum 5072

Query Match 90.2%; Score 37; DB 1; Length 328;
 Best Local Similarity 66.7%; Pred. No. 1.75e+01;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294
 |||:
 Qy 3 PTGEPQ 8

RESULT 9

ENTRY S77299 #type complete
 TITLE C4-dicarboxylase-binding protein - Synecchocystis sp. (strain PCC 6803)
 ALTERNATE_NAMES protein sll1314
 ORGANISM #formal_name Synecchocystis sp.
 #variety PCC 6803
 DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 21-Aug-1998

ACCESSIONS S77299
 REFERENCE S74322

#authors

Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. (1996) 3:109-136

#journal Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.

#cross-references MUID:97061201

#accession S77299
 #status nucleic acid sequence not shown; translation not shown
 #molecule_type DNA

#residues 1-369 #label KAN

#cross-references EMBL:D90907; GB:AB001339; NID:g1652618; PID:d1018366; PID:g1652713

#note the nucleotide sequence was submitted to the EMBL Data Library, June 1996

GENETICS

#gene dctp
 SUMMARY #length 369 #molecular-weight 41164 #checksum 6843

Query Match 90.2%; Score 37; DB 2; Length 369;
 Best Local Similarity 66.7%; Pred. No. 1.75e+01;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 99 PAGDPQ 104
 |||:
 Qy 3 PTGEPQ 8

RESULT 10

ENTRY A36909 #type fragment
 TITLE glutamyl-RNA synthetase homolog gltX - Chlamydia psittaci (fragment)

ORGANISM

#formal_name Chlamydia psittaci

DATE 07-Apr-1994 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998

ACCESSIONS

A36909

```

REFERENCE      A36909
#authors      Wichlan, D.G.; Hatch, T.P.
#journal      J. Bacteriol. (1993) 175:2936-2942
#title        Identification of an early-stage gene of Chlamydia psittaci
              68C.
#cross-references MUID:93259937
#contents     68C
#accession    A36909
#status       preliminary
#molecule_type DNA
#residues     1-370 ##label WIC
#cross-references GB:113598; GB:M73485; NID:g2898332; PID:g289833
#note         sequence extracted from NCBI backbone (NCBIN:131863,
              NCBI:P:131864)
CLASSIFICATION #superfamily glutamate--trna ligase
SUMMARY        #length 370 #checksum 8346

Query Match      90.2%; Score 37; DB 2; Length 370;
Best Local Similarity 66.7%; Pred. No. 1.75e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 16 PTGDPH 21
|||:|
Qy 3 PTGEPQ 8

RESULT 11
ENTRY      S21172      #type complete
TITLE      glutamate--trna ligase (EC 6.1.1.17) - Thermus aquaticus
ORGANISM   glutamate--trna ligase
DATE       22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change
              24-Apr-1998
ACCESSIONS S21172; S21236
REFERENCE   S21172
#authors    Nureki, O.; Suzuki, K.; Hara-Yokoyama, M.; Kohno, T.;
              Matsuzawa, H.; Ohta, T.; Shimizu, T.; Morikawa, K.;
              Miyazawa, T.; Yokoyama, S.
#journal    Eur. J. Biochem. (1992) 204:465-472
#title      GlutamyI-trna synthetase from Thermus thermophilus HB8.
              Molecular cloning of the gltX gene and crystallization of
              the overproduced protein.
#cross-references MUID:92174899
#accession  S21172
#molecule_type DNA
#residues   1-468 ##label NUR
#note       the sequence from Fig. 4 is inconsistent with that from
              Fig. 3 in having 67-Thr
#note       the source is designated as Thermus thermophilus
#accession  S21236
#molecule_type DNA
#residues   1-40 ##label NU2
#note       the source is designated as Thermus thermophilus
CLASSIFICATION #superfamily glutamate--trna ligase
KEYWORDS        aminoacyl-trna synthetase; ligase; protein biosynthesis
SUMMARY        #length 468 #molecular-weight 53901 #checksum 2454

Query Match      90.2%; Score 37; DB 2; Length 468;
Best Local Similarity 66.7%; Pred. No. 1.75e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 10 PTGDPH 15
|||:|
Qy 3 PTGEPQ 8

RESULT 12
ENTRY      S76516      #type complete
TITLE      integral membrane protein - Synecocystis sp. (strain PCC
              6803)
ALTERNATE_NAMES protein slr0642
ORGANISM     #formal_name Synecocystis sp.
#variety     PCC 6803

```

```

DATE          25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
              21-Aug-1998
ACCESSIONS    S76516
REFERENCE     S74322
#authors      Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
              Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugiyama, M.;
              Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
              Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
              S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
              Yasuda, M.; Tabata, S.
#journal      DNA Res. (1996) 3:109-136
#title        Sequence analysis of the genome of the unicellular
              cyanobacterium Synecocystis sp. PCC6803. II. Sequence
              determination of the entire genome and assignment of
              potential protein-coding regions.
#cross-references MUID:97061201
#accession    S76516
#status       nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues     1-494 ##label KAN
#cross-references EMBL:D64002; GB:AB001339; NID:gl001612; PID:dl011013;
              PID:gl001631
#note         the nucleotide sequence was submitted to the EMBL Data
              Library, June 1996
GENETICS      #start_codon GTG
KEYWORDS       membrane protein
SUMMARY        #length 494 #molecular-weight 53585 #checksum 7038

Query Match      90.2%; Score 37; DB 2; Length 494;
Best Local Similarity 66.7%; Pred. No. 1.75e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 437 PAGDPQ 442
|||:|
Qy 3 PTGEPQ 8

RESULT 13
ENTRY      JC5208      #type complete
TITLE      glutamate--trna ligase (EC 6.1.1.17) - Chlamydia psittaci
ALTERNATE_NAMES glutamyl-trna synthetase
ORGANISM        #formal_name Chlamydia psittaci
DATE            20-Feb-1997 #sequence_revision 27-Feb-1997 #text_change
              14-Nov-1997
ACCESSIONS    JC5208
REFERENCE     Hsia, R.; Bavoil, P.M.
#authors      Gene (1996) 176:163-169
#journal      Homologs of Escherichia coli recJ, gltX and of a putative
#title        'early' gene of avian Chlamydia psittaci are located
              upstream of the 'late' omp2 locus of Chlamydia psittaci
              strain guinea pig inclusion conjunctivitis.
#accession    JC5208
#status       preliminary
#molecule_type DNA
#residues     1-505 ##label HSI
#cross-references GB:U41759; NID:gl783376; PID:gl783380
#experimental_source strain GPIC
CLASSIFICATION #superfamily glutamate--trna ligase
KEYWORDS        aminoacyl-trna synthetase; ligase; protein biosynthesis
FEATURE        9-11
              322
              66
SUMMARY        #length 505 #molecular-weight 58544 #checksum 9458
              #region ATP binding #status predicted\
              #region trna binding #status predicted\
              #binding_site glutamate (trp) #status predicted\

Query Match      90.2%; Score 37; DB 2; Length 505;
Best Local Similarity 66.7%; Pred. No. 1.75e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 14 PTGDPH 19
|||:|

```

QY 3 PTGEPQ 8

RESULT 14
ENTRY F71513 #type complete
TITLE Probable glutamyl-tRNA synthetase - Chlamydia trachomatis
(serotype D, strain UW3/Cx)
ORGANISM #formal_name Chlamydia trachomatis
DATE 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change
13-Sep-1998
ACCESSIONS F71513
REFERENCE A71460
#authors Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe,
R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov,
R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W.
#submission submitted to GenBank, May 1998
#description Genome sequence of an obligate intracellular pathogen of
humans: Chlamydia trachomatis.
#accession F71513
##status preliminary
##molecule_type DNA
##residues 1-506 ##label ARN
##cross-references GB:AE001318; GB:AE001273; NID:g3328875; PID:g3328877
##experimental_source serotype D, strain UW-3/Cx

GENETICS
#gene gltX
SUMMARY #length 506 #molecular-weight 58510 #checksum 7203

Query Match 90.2%; Score 37; DB 2; Length 506;
Best Local Similarity 66.7%; Pred. No. 1.75e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 14 PTGDPH 19

QY 3 PTGEPQ 8

RESULT 15
ENTRY SKPSXT #type complete
TITLE secretion protein xcpT - Pseudomonas aeruginosa
ORGANISM #formal_name Pseudomonas aeruginosa
DATE 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change
05-Sep-1997
ACCESSIONS S25386; S22731
REFERENCE S25384
#authors Bally, M.; Filloux, A.; Akrim, M.; Ball, G.; Lazdunski, A.;
Tomassen, J.
#journal Mol. Microbiol. (1992) 6:1121-1131
#title Protein secretion in Pseudomonas aeruginosa: characterization
of seven xcp genes and processing of secretory apparatus
components by prepilin peptidase.
#accession S25386
##molecule_type DNA
##residues 1-142 ##label BAL
##cross-references EMBL:X62666; NID:g45433; PID:g581440
GENETICS
#gene xcpT
#start_codon TTG
CLASSIFICATION #superfamily secretion protein xcpT
SUMMARY #length 142 #molecular-weight 15449 #checksum 2410

Query Match 87.8%; Score 36; DB 1; Length 142;
Best Local Similarity 83.3%; Pred. No. 3.05e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 79 PTGNPQ 84

QY 3 PTGEPQ 8

Search completed: Thu May 13 15:40:47 1999
Job time : 10 secs.

M P E R F

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPerch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu May 13 15:36:54 1999; MasPar time 2.16 Seconds
Tabular output not generated. 124.372 Million cell updates/sec

Title: >US-09-040-485-8
Description: (1-10) from US09040485.pep (1 of 2)
Perfect Score: 67
Sequence: 1 DGPTEGPQEE 10
Scoring table: PAM 150
Gap 15
Searched: 74019 seqs, 26840295 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: swiss-prot35
1:swissprot
Statistics: Mean 21.176; Variance 23.637; scale 0.896

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	67	100.0	757	1	ASPH_HUMAN	1.38e-05
2	50	74.6	232	1	VG08_HSVI1	3.41e-01
3	47	70.1	433	1	TRBI_AGR6	1.72e+00
4	47	70.1	1389	1	TIMDROME	1.72e+00
5	47	70.1	3567	1	ERY2_SACER	1.72e+00
6	46	68.7	311	1	SRY_MUSST	2.90e+00
7	46	68.7	355	1	SEX-DETERMINING REGION	2.90e+00
8	45	67.2	53	1	CALL1_RABIT	2.90e+00
9	45	67.2	259	1	YCBCECOLI	4.85e+00
10	45	67.2	1027	1	P531_HUMAN	4.85e+00
11	45	67.2	1418	1	CALL2_HUMAN	4.85e+00
12	45	67.2	1596	1	MAM1_DROME	4.85e+00
13	44	65.7	261	1	PRP2_MOUSE	8.07e+00
14	44	65.7	745	1	ALD_HUMAN	8.07e+00
15	44	65.7	755	1	RREL_HUMAN	8.07e+00
16	44	65.7	1388	1	CALL1_HUMAN	8.07e+00
17	43	64.2	336	1	CALL6_HUMAN	8.07e+00
18	43	64.2	449	1	CMGA_BOVIN	1.33e+01
19	43	64.2	1670	1	CA34_HUMAN	1.33e+01
20	43	64.2	1707	1	CA24_MOUSE	1.33e+01
21	43	64.2	1841	1	RPB1_ARATH	1.33e+01
22	43	64.2	1860	1	RPB0_ARATH	1.33e+01
23	42	62.7	181	1	HP28_RAT	2.17e+01

24	42	62.7	181	1	HP28_HUMAN	2.17e+01
25	42	62.7	218	1	POU2_XENLA	2.17e+01
26	42	62.7	268	1	EP34_HCMVA	2.17e+01
27	42	62.7	333	1	CAP1_DICDI	2.17e+01
28	42	62.7	339	1	YG2J_YEAST	2.17e+01
29	42	62.7	384	1	COG1_RANCA	2.17e+01
30	42	62.7	501	1	IRK3_HUMAN	2.17e+01
31	42	62.7	501	1	Y225_MYCTU	2.17e+01
32	42	62.7	535	1	Y225_MYCTU	2.17e+01
33	42	62.7	674	1	CALA_BOVIN	2.17e+01
34	42	62.7	685	1	YICA_BACSU	2.17e+01
35	42	62.7	736	1	ALD_MOUSE	2.17e+01
36	42	62.7	760	1	ALK1_YEAST	2.17e+01
37	42	62.7	991	1	BMPI_MOUSE	2.17e+01
38	42	62.7	1019	1	CA16_CHICK	2.17e+01
39	42	62.7	1027	1	CAFF_RIFPA	2.17e+01
40	42	62.7	1063	1	ELF1_DROME	2.17e+01
41	42	62.7	1365	1	GTFS_STRDO	2.17e+01
42	42	62.7	1459	1	CA12_MOUSE	2.17e+01
43	42	62.7	1496	1	CA25_HUMAN	2.17e+01
44	42	62.7	2336	1	CIC5_RAT	2.17e+01
45	42	62.7	2339	1	CIC5_HUMAN	2.17e+01
			2541	1	TALI_MOUSE	2.17e+01

ALIGNMENTS

RESULT 1
ID ASPH_HUMAN STANDARD; PRT; 757 AA.
AC Q12797;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-HYDROXYLASE).
DE DIOXYGENASE).
GN ASPH.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95121937.
RA KORIOTH F., GIEFFERS C., FREY J.;
RL GENE 1501395-15911994;
CC -!- FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF PROTEINS.
CC -!- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) -> PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).
CC -!- COFACTOR: IRON.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM.
CC -!- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES TESTED.
CC -!- PFM: MIGHT BE PROCESSED TO THE 56 KD (AA 274-757) OR 52 KD (AA 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY SIMILARITY).
CC EMBL: U03109; E82591; .
DR MIM: 600582; .
KW OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR; ENDOPLASMIC RETICULUM.
FT DOMAIN 1 54
FT TRANSMEM 55 75
FT DOMAIN 76 757
FT DOMAIN 13 20
FT DOMAIN 323 332
FT CARBOHYD 452 452
FT CARBOHYD 705 705
SQ SEQUENCE 757 AA; 85498 MW; AE6AFC24 CRC32;

Query Match 100.0%; Score 67; DB 1; Length 757;
Best Local Similarity 100.0%; Pred. No. 1.38e-05;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 173 DGTPGEPOQ 182
|||||
QY 1 DGTPGEPOQ 10

RESULT 2
ID VG08_HSV11 STANDARD; PRT; 232 AA.

AC 000137;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL GENE 8 MEMBRANE PROTEIN.
GN 8.

OS ICTALURID HERPESVIRUS 1 (CHANNEL CATEFISH VIRUS) (CCV).

OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-AUBURN 1.

RX MEDLINE; 92087490.

RA DAVISON A.J.;

RL VIROLOGY 186:9-14(1992).

DR EMBL; M75136; G331295; -.

DR EMBL; M75136; G331218; -.

DR PIR; I36786; MMBE13

KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.

FT TRANSMEM 166 182 POTENTIAL.

FT TRANSMEM 195 211 POTENTIAL.

SQ SEQUENCE 232 AA; 25466 MW; AA90E31E CRC32;

Query Match 74.6%; Score 50; DB 1; Length 232;

Best Local Similarity 50.0%; Pred. No. 3.41e-01;

Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 88 ESPGEPHRD 97
:|||||
QY 1 DGTPGEPOQ 10

RESULT 3

ID TRB1_AGR6 STANDARD; PRT; 433 AA.

AC P54917;

DT 01-OCT-1996 (REL. 34, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE CONJUGAL TRANSFER PROTEIN TRB1.

GN TRB1.

OS AGROBACTERIUM TUMEFACIENS.

OG PLASMID PTIAGNC.

OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;

OC RHIZOBIACEAE.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 96312368.

RA ALT-MORBE J.; STRIKER J.L.; FUQUA C.; LI P.L.; FARRAND S.K.;

RA WINANS S.C.;

RL J. BACTERIOL. 178:4248-4257(1996).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

DR EMBL; U43675; G2749901; -.

KW CONJUGATION; PLASMID; TRANSMEMBRANE.

FT TRANSMEM 25 45 POTENTIAL.

FT TRANSMEM 307 327 POTENTIAL.

FT TRANSMEM 347 367 POTENTIAL.

SQ SEQUENCE 433 AA; 47258 MW; E3C35AE6 CRC32;

Query Match 70.1%; Score 47; DB 1; Length 433;

Best Local Similarity 70.0%; Pred. No. 1.72e+00;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 79 DGIIGEPQOQ 88
|||||
QY 1 DGTPGEPOQ 10

RESULT 4

ID TIM_DRONE STANDARD; PRT; 1389 AA.

AC P49021;

DT 01-FEB-1996 (REL. 33, CREATED)

DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)

DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)

DE TIMELESS PROTEIN.

GN TIM.

OS DROSOPHILA MELANOGASTER (FRUIT FLY).

OC EURKYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 96055118.

RA MYERS M.P.; WAGER-SMITH K.; WESLEY C.S.; YOUNG M.W.; SEHGAL A.;

RL SCIENCE 270:805-808(1995).

RN [2]

RP INTERACTION WITH PER.

RX MEDLINE; 96055120.

RA GERAKIS N.; SAEZ L.; DELAHAYE-BROWN A.M.; MYERS M.P.; SEHGAL A.;

RA YOUNG M.W.; WEITZ C.J.;

RL SCIENCE 270:811-815(1995).

CC -1- FUNCTION: REQUIRED FOR THE PRODUCTION OF CIRCADIAN RHYTHMS.

CC INTERACTS WITH PERIOD (PER). MAY BE REQUIRED AT A SPECIFIC

CC TIME OF DAY TO ALLOW ACCUMULATION AND NUCLEAR LOCALIZATION OF

CC THE PER PROTEIN.

DR EMBL; U37018; G1050970; -.

DR FLYBASE; FBgn0014396; tim.

KW BIOLOGICAL RHYTHMS.

FT DOMAIN 351 380 ASP/GLU-RICH (ACIDIC).

FT DOMAIN 340 353 ARG/LYS-RICH (BASIC).

SQ SEQUENCE 1389 AA; 155665 MW; 1DB78941 CRC32;

Query Match 70.1%; Score 47; DB 1; Length 1389;

Best Local Similarity 60.0%; Pred. No. 1.72e+00;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 510 DGPGKPGHQ 519
||| |
QY 1 DGTPGEPOQ 10

RESULT 5

ID ERY2_SACER STANDARD; PRT; 3567 AA.

AC Q03132; Q54096;

DT 01-OCT-1993 (REL. 27, CREATED)

DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE ERYTHRONOLIDE SYNTHASE, MODULES 3 AND 4 (EC 2.3.1.94) (ORF 2) (6-DEOXYERYTHRONOLIDE B SYNTHASE II) (DEBS 2).

GN ERYA.

OS SACCHAROPOLYSPORA ERYTHRAEA (STREPTOMYCES ERYTHRAEUS).

OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; NOCARDIOFORM.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 91220065.

RA DONADIO S.; STAVER M.J.; MCALPINE J.B.; SWANSON S.J.; KATZ L.;

RL SCIENCE 252:675-679(1991).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-NRRL 2338;

RX MEDLINE; 92155230.

RA BEVITT D.J.; CORTES J.; HAYDOCK S.F.; LEADLAY P.F.;

RL EUR. J. BIOCHEM. 204:39-49(1992).

CC -1- CATALYTIC ACTIVITY: 6 METHYLMALONYL-COA + PROPIONYL-COA -> 7 COA

CC + 6-DEOXYERYTHRONOLIDE B.

CC -1- COFACTOR: NADP.

CC -1- COFACTOR: CONTAINS TWO COVALENTLY BOUND PHOSPHOPANTHETINES.

CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN

CC BIOSYNTHESIS.

CC -1- IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH ENCODING FOR A

CC FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3 ORFS CODES

FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5, AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6, RESPECTIVELY.

-1- BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT), BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR), DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE BETA CARBON, AND AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION OF THE FULL-LENGTH CHAIN.

-1- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).

EMBL: M63677; G152694; -.

EMBL: X62569; G581651; -.

DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 2.

DR PROSITE: PS00606; B-KETOACYL SYNTHASE; 2.

DR PROSITE: PS50075; ACP_DOMAIN; 2.

KW TRANSFERASE: ACYLTRANSFERASE; ANTI-BIOTIC BIOSYNTHESIS; NADP;

KW PHOSPHOPANTHETHEINE; MULTIFUNCTIONAL ENZYME.

FT DOMAIN 1 1484 MODULE 3.

FT DOMAIN 1485 3567 MODULE 4.

FT DOMAIN 27 488 BETA-KETOACYL SYNTHASE (KS).

FT DOMAIN 559 884 ACYLTRANSFERASE (AT).

FT DOMAIN 1130 1301 BETA-KETOACYL REDUCTASE (KR) (POSSIBLY NON-FUNCTIONAL).

FT DOMAIN 1397 1467 ACYL CARRIER (ACP).

FT DOMAIN 1485 1943 BETA-KETOACYL SYNTHASE (KS).

FT DOMAIN 2013 2336 ACYLTRANSFERASE (AT).

FT DOMAIN 2383 3066 DEHYDRATASE/ENOYLREDUCTASE (DH/ER).

FT DOMAIN 3139 3322 BETA-KETOACYL REDUCTASE (KR).

FT DOMAIN 3415 3485 ACYL CARRIER (ACP).

FT ACT_SITE 202 202 THIOESTER BOND.

FT ACT_SITE 651 651 ACYL-ENZYME INTERMEDIATE.

FT BINDING 1430 1430 PHOSPHOPANTHETHEINE (BY SIMILARITY).

FT ACT_SITE 1661 1661 THIOESTER BOND.

FT ACT_SITE 2115 2115 ACYL-ENZYME INTERMEDIATE.

FT NP_BIND 2961 2978 NADP (ER).

FT NP_BIND 3142 3157 NADP (KR).

FT BINDING 3448 3448 PHOSPHOPANTHETHEINE (BY SIMILARITY).

FT CONFLICT 438 438 R -> A (IN REF. 2).

FT CONFLICT 480 480 T -> S (IN REF. 2).

FT CONFLICT 1241 1241 L -> F (IN REF. 2).

FT CONFLICT 2664 2664 G -> V (IN REF. 2).

SQ SEQUENCE 3567 AA; 374413 MW; A2F5EA2C CRC32;

Query Match 70.1%; Score 47; DB 1; Length 3567;
Best Local Similarity 87.5%; Pred. No. 1.72e+00;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1376 GPTGEPAQ 1383
QY 2 GPTGEPAQ 9
| | | | | | | |

RESULT 6 STANDARD; PRT; 311 AA.

ID SRY_MUSSI
AC Q62565;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE SEX-DETERMINING REGION Y PROTEIN (TESTIS-DETERMINING FACTOR).
GN SRY OR TDY.
OS MUS SPICILEGUS (STEPPE MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98043417.
RA ALBRECHT K.H., EICHER E.M.;
RL GENETICS 147:1267-1277(1997).
RN [2]
RP SEQUENCE OF 1-143 FROM N.A.
RX MEDLINE; 93361118.
RA TUCKER P.K., LUNDRIGAN B.L.;
RL NATURE 364:715-717(1993).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR WHICH REGULATES A GENETIC SWITCH IN MALE DEVELOPMENT. IT IS RESPONSIBLE FOR INITIATING MALE SEX DETERMINATION. SRY HMG BOX RECOGNIZES DNA BY PARTIAL INTERCALATION IN THE MINOR GROOVE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DOMAIN: THE GLN- AND HIS-RICH DOMAIN MAY MEDIATE PROTEIN-PROTEIN INTERACTIONS.
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
DR EMBL; U70659; G2623383; -.
DR EMBL; AF009521; G2271483; -.
DR EMBL; L29544; G496150; -.
DR MGD; MGI:988660; TDY.
KW DNA-BINDING; NUCLEAR PROTEIN; TRANSCRIPTION REGULATION; ACTIVATOR;
KW SEXUAL DIFFERENTIATION; REPEAT.
FT DNA_BIND 5 73 HMG BOX.
FT VARIANT 274 274 K -> R.
SQ SEQUENCE 355 AA; 44260 MW; 09CSACAE CRC32;

Query Match 68.7%; Score 46; DB 1; Length 355;
Best Local Similarity 70.0%; Pred. No. 2.90e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 137 DIPTGHPQQ 146
QY 1 DGPTGEPAQ 10
| | | | | | | |

RESULT 7 STANDARD; PRT; 355 AA.

ID SRY_MUSSP
AC Q62563;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE SEX-DETERMINING REGION Y PROTEIN (TESTIS-DETERMINING FACTOR).
GN SRY.
OS MUS SPRETUS (WESTERN WILD MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98043417.
RA ALBRECHT K.H., EICHER E.M.;
RL GENETICS 147:1267-1277(1997).
RN [2]
RP SEQUENCE OF 1-143 FROM N.A.
RX MEDLINE; 93361118.
RA TUCKER P.K., LUNDRIGAN B.L.;
RL NATURE 364:715-717(1993).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR WHICH REGULATES A GENETIC SWITCH IN MALE DEVELOPMENT. IT IS RESPONSIBLE FOR INITIATING MALE SEX DETERMINATION. SRY HMG BOX RECOGNIZES DNA BY PARTIAL INTERCALATION IN THE MINOR GROOVE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DOMAIN: THE GLN- AND HIS-RICH DOMAIN MAY MEDIATE PROTEIN-PROTEIN INTERACTIONS.
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
DR EMBL; U70659; G2623383; -.
DR EMBL; AF009521; G2271483; -.
DR EMBL; L29544; G496150; -.
DR MGD; MGI:988660; TDY.
KW DNA-BINDING; NUCLEAR PROTEIN; TRANSCRIPTION REGULATION; ACTIVATOR;
KW SEXUAL DIFFERENTIATION; REPEAT.
FT DNA_BIND 5 73 HMG BOX.
FT VARIANT 274 274 K -> R.
SQ SEQUENCE 355 AA; 44260 MW; 09CSACAE CRC32;

Query Match 68.7%; Score 46; DB 1; Length 355;
Best Local Similarity 70.0%; Pred. No. 2.90e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 137 DIPTGHPQQ 146
QY 1 DGPTGEPAQ 10
| | | | | | | |

RA LUNDRIGAN B.L., TUCKER P.K.;
RL MOL. BIOL. EVOL. 11:483-492(1994).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR WHICH REGULATES A GENETIC SWITCH IN MALE DEVELOPMENT. IT IS RESPONSIBLE FOR INITIATING MALE SEX DETERMINATION. SRY HMG BOX RECOGNIZES DNA BY PARTIAL INTERCALATION IN THE MINOR GROOVE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DOMAIN: THE GLN- AND HIS-RICH DOMAIN MAY MEDIATE PROTEIN-PROTEIN INTERACTIONS.
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
DR EMBL; U70658; G2623381; -.
DR EMBL; AF009520; G2271481; -.
DR EMBL; L29550; G463146; -.
DR MGD; MGI:988660; TDY.
KW DNA-BINDING; NUCLEAR PROTEIN; TRANSCRIPTION REGULATION; ACTIVATOR;
KW SEXUAL DIFFERENTIATION; REPEAT.
FT DNA_BIND 5 73 HMG BOX.
SQ SEQUENCE 311 AA; 38469 MW; 035872C0 CRC32;

Query Match 68.7%; Score 46; DB 1; Length 311;
Best Local Similarity 70.0%; Pred. No. 2.90e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 137 DIPTGHPQQ 146
QY 1 DGPTGEPAQ 10
| | | | | | | |

RESULT 7 STANDARD; PRT; 355 AA.

ID SRY_MUSSP
AC Q62563;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE SEX-DETERMINING REGION Y PROTEIN (TESTIS-DETERMINING FACTOR).
GN SRY.
OS MUS SPRETUS (WESTERN WILD MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98043417.
RA ALBRECHT K.H., EICHER E.M.;
RL GENETICS 147:1267-1277(1997).
RN [2]
RP SEQUENCE OF 1-143 FROM N.A.
RX MEDLINE; 93361118.
RA TUCKER P.K., LUNDRIGAN B.L.;
RL NATURE 364:715-717(1993).
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR WHICH REGULATES A GENETIC SWITCH IN MALE DEVELOPMENT. IT IS RESPONSIBLE FOR INITIATING MALE SEX DETERMINATION. SRY HMG BOX RECOGNIZES DNA BY PARTIAL INTERCALATION IN THE MINOR GROOVE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DOMAIN: THE GLN- AND HIS-RICH DOMAIN MAY MEDIATE PROTEIN-PROTEIN INTERACTIONS.
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.
DR EMBL; U70659; G2623383; -.
DR EMBL; AF009521; G2271483; -.
DR EMBL; L29544; G496150; -.
DR MGD; MGI:988660; TDY.
KW DNA-BINDING; NUCLEAR PROTEIN; TRANSCRIPTION REGULATION; ACTIVATOR;
KW SEXUAL DIFFERENTIATION; REPEAT.
FT DNA_BIND 5 73 HMG BOX.
FT VARIANT 274 274 K -> R.
SQ SEQUENCE 355 AA; 44260 MW; 09CSACAE CRC32;

Query Match 68.7%; Score 46; DB 1; Length 355;
Best Local Similarity 70.0%; Pred. No. 2.90e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 137 DIPTGHPQQ 146
QY 1 DGPTGEPAQ 10
| | | | | | | |

```

QY 1 DGPTGEPQOE 10

RESULT 8
ID CALL_RABBIT STANDARD; PRT; 53 AA.
AC P02456;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE COLLAGEN ALPHA 1(I) CHAIN (FRAGMENT).
GN COL1A1.
OS ORYCTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; LAGOMORPHA.
RN [1]
RP SEQUENCE.
RX MEDLINE; 70252720.
RA BORNSTEIN P., NESSE R.;
RL ARCH. BIOCHEM. BIOPHYS. 138:443-450(1970).
CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC (FIBRILLAR FORMING COLLAGEN).
CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC HYDROXYAPATITE.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
DR PIR; A02856; CGBR15.
DR HSP; P12108; IBBF.
DR PROSITE; PS01208; VWFIC; PARTIAL.
KW EXTRACELLULAR MATRIX; CONNECTIVE TISSUE; REPEAT; HYDROXYLATION;
KW COLLAGEN.
FT MOD_RES 7 7 CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT INVOLVED IN CROSS-LINKING.
FT MOD_RES 26 26 HYDROXYLATION (PROBABLE).
FT MOD_RES 29 29 HYDROXYLATION (PROBABLE).
FT MOD_RES 32 32 HYDROXYLATION (PROBABLE).
FT MOD_RES 41 41 HYDROXYLATION (PROBABLE).
FT MOD_RES 44 44 HYDROXYLATION (PROBABLE).
FT MOD_RES 47 47 HYDROXYLATION (PROBABLE).
FT MOD_RES 53 53 HYDROXYLATION (PROBABLE).
FT NON_TER 53
SQ SEQUENCE 53 AA; 4987 MW; B20D776E CRC32;

Query Match 67.2%; Score 45; DB 1; Length 53;
Best Local Similarity 44.4%; Pred. No. 4.85e+00;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 38 ZGPGZPGZ 46
QY 1 DGPTGEPQOE 9

RESULT 9
ID YCBC_ECOLI STANDARD; PRT; 259 AA.
AC P36565; P75846;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 28.7 KD PROTEIN IN KDSB-MUKF INTERGENIC REGION.
GN YCBC.
OS ESCHERICHIA COLI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RA BLATTNER F.R., PLUNKETT G. III, MAYHEW G.F., PERNA N.T., GLASNER F.D.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RN SEQUENCE OF 1-170 FROM N.A.
RC STRAIN-K12 / W3110;
RX MEDLINE; 94232180.
RA FENG J., YAMANAKA K., NIKI H., OGURA T., HIRAGA S.;

QY 1 DGPTGEPQOE 10

RESULT 11
ID CALL_HUMAN STANDARD; PRT; 1418 AA.
AC P02458;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROCOLLAGEN ALPHA 1(II) CHAIN PRECURSOR (CONTAINS: CHONDROCALCIN).
GN COL2A1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90067946.
RA SU M.W., LEE B., RAMIREZ F., MACHADO M., HORTON W.;
RL NUCLEIC ACIDS RES. 17:9473-9473(1989).
RN [2]
RN SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE; 87031574.
RA NUNEZ A.M., KOHNO K., MARTIN G.R., YAMADA Y.;
RL GENE 44:11-16(1986).
RN [3]

RL MOL. GEN. GENET. 243:136-147(1994).
DR EMBL; AF000194; G1787150; -.
DR EMBL; D26440; -. NOT_ANNOTATED_CDS.
DR ECOGENE; EG12166; YCBC.
KW HYPOTHETICAL PROTEIN.
FT CONFLICT 89 89 Y -> N (IN REF. 2).
FT CONFLICT 150 171 GVPRLITLDPKTEEEAAA -> ACRASKLSPWICQKI
FT PAKKIQ (IN REF. 2).
SQ SEQUENCE 259 AA; 28666 MW; CB9EFE64 CRC32;

Query Match 67.2%; Score 45; DB 1; Length 259;
Best Local Similarity 66.7%; Pred. No. 4.85e+00;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 251 GSGGEPRQE 259
QY 2 GPTGEPQOE 10

RESULT 10
ID P531_HUMAN STANDARD; PRT; 1027 AA.
AC Q12888;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE P53-BINDING PROTEIN 53BPI (FRAGMENT).
GN TP53BP1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94286584.
RA IWABUCHI K., BARTEL P.L., LI B., MARRACCINO R., FIELDS S.;
RL PROC. NATL. ACAD. SCI. U.S.A. 91:6098-6102(1994).
CC -1- FUNCTION: BINDS TO THE CENTRAL DOMAIN OF P53.
DR EMBL; U09477; G488592; -.
FT NON_TER 1 1
FT DOMAIN 697 701 POLY-SER.
FT DOMAIN 815 819 POLY-GLU.
FT SEQUENCE 1027 AA; 111134 MW; 06D3FABB CRC32;

Query Match 67.2%; Score 45; DB 1; Length 1027;
Best Local Similarity 60.0%; Pred. No. 4.85e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 808 DGPTGSSEE 817
QY 1 DGPTGEPQOE 10

RESULT 11
ID CALL_HUMAN STANDARD; PRT; 1418 AA.
AC P02458;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROCOLLAGEN ALPHA 1(II) CHAIN PRECURSOR (CONTAINS: CHONDROCALCIN).
GN COL2A1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90067946.
RA SU M.W., LEE B., RAMIREZ F., MACHADO M., HORTON W.;
RL NUCLEIC ACIDS RES. 17:9473-9473(1989).
RN [2]
RN SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE; 87031574.
RA NUNEZ A.M., KOHNO K., MARTIN G.R., YAMADA Y.;
RL GENE 44:11-16(1986).
RN [3]

```

RP SEQUENCE OF 432-1145 FROM N.A.
RA RAMIREZ F.;
RL SUBMITTED (DEC-1988) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE OF 963-1418 FROM N.A.
RX MEDLINE; 85190334.
RA CHEAH K.S.E., STOKER N.G., GRIFFIN J.R., GROSVELD F.G., SOLOMON E.;
RL PROC. NATL. ACAD. SCI. U.S.A. 82:2555-2559(1985).
RN [5]
RP SEQUENCE OF 1120-1398 FROM N.A.
RX MEDLINE; 85306861.
RA ELIMA K., MAEKELAE J.K., VUORIO T., KAUPPINEN S., KNOWLES J.,
RA VUORIO E.;
RL BIOCHEM. J. 229:183-188(1985).
RN [6]
RP SEQUENCE OF 1106-1418 FROM N.A.
RX MEDLINE; 88067771.
RA ELIMA K., VUORIO T., VUORIO E.;
RL NUCLEIC ACIDS RES. 15:9499-9504(1987).
RN [7]
RP SEQUENCE OF 1227-1289 FROM N.A.
RX MEDLINE; 86104139.
RA NUNEZ A.M., FRANCOMANO C., YOUNG M.F., MARTIN G.R., YAMADA Y.;
RL BIOCHEMISTRY 24:6343-6348(1985).
RN [8]
RP SEQUENCE OF 1176-1226 FROM N.A.
RX MEDLINE; 84118798.
RA STROM C.M., UPHOLT W.B.;
RL NUCLEIC ACIDS RES. 12:1025-1038(1984).
RN [9]
RP SEQUENCE OF 35-167 FROM N.A.
RX MEDLINE; 89233138.
RA SU M.W., BENSON-CHANDA V., VISSING H., RAMIREZ F.;
RL GENOMICS 4:438-441(1989).
RN [10]
RP REVIEW ON VARIANTS.
RX MEDLINE; 91184577.
RA KUIVANEMI H., TROMP G., PROCKOP D.J.;
RL FASEB J. 5:2052-2060(1991).
RN [11]
RP REVIEW ON VARIANTS.
RX MEDLINE; 97255959.
RA KUIVANEMI H., TROMP G., PROCKOP D.J.;
RL HUM. MUTAT. 9:300-315(1997).
RN [12]
RP VARIANT SER-1074.
RX MEDLINE; 90036909.
RA VISSING H., D'ALESSIO M., LEE B., RAMIREZ F., GODFREY M.,
RA HOLLISTER D.W.;
RL J. BIOL. CHEM. 264:18265-18267(1989).
RN [13]
RP VARIANT SEDC GLY-1095--TYR-1330 DEL.
RX MEDLINE; 89266907.
RA LEE B., VISSING H., RAMIREZ F., ROGERS D., RIMOIN D.;
RL SCIENCE 244:978-980(1989).
RN [14]
RP VARIANT OSTEOARTHRITIS CYS-650.
RX MEDLINE; 90370826.
RA ALA-KOKKO L., BALDWIN C.T., MOSKOWITZ R.W., PROCKOP D.J.;
RL PROC. NATL. ACAD. SCI. U.S.A. 87:6565-6568(1990).
RN [15]
RP VARIANT OI-IV VAL-717.
RX MEDLINE; 91291136.
RA BATEMAN J.F., HANNAGAN M., CHAN D., COLE W.G.;
RL BIOCHEM. J. 276:765-770(1991).
RN [16]
RP VARIANT OSTEOARTHRITIS CYS-650.
RX MEDLINE; 91086471.
RA EYRE D.R., WEIS M.A., MOSKOWITZ R.W.;
RL J. CLIN. INVEST. 87:357-361(1991).
RN [17]
RP VARIANT HYPOCHONDROGENESIS GLU-984.
RX MEDLINE; 93054548.

RA BOGAERT R., TILLER G.E., WIES M.A., GRUBER H.E., RIMOIN D.L.,
RA COHN D.H., EYRE D.R.;
RL J. BIOL. CHEM. 267:22522-22526(1992).
RN [18]
RP VARIANT HYPOCHONDROGENESIS SER-705.
RX MEDLINE; 92262484.
RA HORTON W.A., MACHADO M.A., ELLARD J., CAMPBELL D., BARTLEY J.,
RA RAMIREZ F., VITALE E., LEE B.;
RL PROC. NATL. ACAD. SCI. U.S.A. 89:4583-4587(1992).
RN [19]
RP VARIANT WS-II ASP-198.
RX MEDLINE; 93304428.
RA KORKKO J., RITVANEMI P., HAATAJA L., KAARIAINEN H., KIVIRIKKO K.I.,
RA PROCKOP D.J., ALA-KOKKO L.;
RL AM. J. HUM. GENET. 53:55-61(1993).
RN [20]
RP VARIANT SEMD CYS-940.
RA TILLER G.E., WEIS M.A., LACHMAN R.S., COHN D.H., RIMOIN D.L.,
RA EYRE D.R.;
RL AM. J. HUM. GENET. 53:209-209(1993).
RN [21]
RP VARIANT OSTEOARTHRITIS CYS-650.
RX MEDLINE; 93282819.
RA HODERAUM D., MALEMUD C.J., MOSKOWITZ R.W., HAQOI T.M.;
RL BIOCHEM. BIOPHYS. RES. COMMUN. 192:1169-1174(1993).
RN [22]
RP VARIANT SEMD ARG-285.
RX MEDLINE; 93252400.
RA VIRKULA M., RITVANEMI P., VUORIO A.F., KAITILA I., ALA-KOKKO L.,
RA PELTONEN L.;
RL GENOMICS 16:282-285(1993).
RN [23]
RP VARIANT SEDC CYS-206.
RX MEDLINE; 94063862.
RA WILLIAMS C.J., CONSIDINE E.L., KNOWLTON R.G., REGINATO A., NEUMANN G.,
RA HARRISON D., BUXTON P., JIMENEZ S., PROCKOP D.J.;
RL HUM. GENET. 92:499-505(1993).
RN [24]
RP VARIANT SEDC CYS-920.
RX MEDLINE; 93315508.
RA CHAN D., TAYLOR T.K.F., COLE W.G.;
RL J. BIOL. CHEM. 268:15238-15245(1993).
RN [25]
RP VARIANT SEDC SER-1128.
RX MEDLINE; 93140139.
RA COLE W.G., HALL R.K., ROGERS J.G.;
RL J. MED. GENET. 30:27-35(1993).
RN [26]
RP VARIANT AOM ALA-233--LYS-239 DEL.
RX MEDLINE; 95067975.
RA BOGAERT R., WILKIN D., WILCOX W.R., LACHMAN R., RIMOIN D., COHN D.H.,
RA EYRE D.R.;
RL AM. J. HUM. GENET. 55:1128-1136(1994).
RN [27]
RP VARIANT KS ASP-234.
RX MEDLINE; 95179117.
RA WILKIN D.J., BOGAERT R., LACHMAN R.S., RIMOIN D.L., EYRES D.R.,
RA COHN D.H.;
RL HUM. MOL. GENET. 3:1999-2003(1994).
RN [28]
RP VARIANT OI-III VAL-717.
RX MEDLINE; 95187161.
RA FORLINO A., ZOLEZZI F., VALLI M., PIGNATTI P.F., CETTA G.,
RA BRUNELLI P.C., MOTTES M.;
RL HUM. MOL. GENET. 3:2201-2206(1994).
RN [29]
RP VARIANT SEDC SER-378.
RX MEDLINE; 94290498.
RA RITVANEMI P., SOKOLOV B.P., WILLIAMS C.J., CONSIDINE W., YURGNEV L.,
RA MEERSON E.M., ALA-KOKKO L., PROCKOP D.J.;
RL HUM. MUTAT. 3:261-267(1994).
RN [30]
RP VARIANTS CYS-206; CYS-650; ARG-822; SER-1107 AND ARG-1119.

RX MEDLINE; 95276749.
 RA WILLIAMS C.J., ROCK M., CONSIDINE E., MCCARRON S., GOW P., LADDA R.,
 RA MCCLAIN D., MICHELS V.M., MURPHY W., PROCKOP D.J., GANGULY A.,
 RL HUM. MOL. GENET. 4:309-312(1995).
 RN [31]
 RP VARIANT HYPOCHONDROGENESIS CYS-1044.
 RX MEDLINE; 96298263.
 RA MUNDLOS S., CHAN D., MCGILL J., BATEMAN J.F.,
 RL AM. J. MED. GENET. 63:129-136(1996).
 RN [32]
 RP VARIANT KS PRO-1138--GLY-1143 DEL.
 RX MEDLINE; 97016530.
 RA WINTERPACHT A., SUPERTI-FURGA A., SCHWARZE U., STOESE H.,
 RA STEINMANN B., SPRANGER J., ZABEL B.,
 RL J. MED. GENET. 33:649-654(1996).
 CC -!- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.
 CC -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -!- DISEASE: DEFECTS IN COL2A1 ARE THE CAUSE OF A VARIETY OF
 CC CHONDRODYSPLASIA INCLUDING HYPOCHONDROGENESIS AND OSTEOARTHRITIS.
 CC -!- DISEASE: DEFECTS IN COL2A1 ARE THE CAUSE OF SPONDYLOEPHYSAL
 CC DYSPLASIA, CONGENITAL TYPE (SDC); A DISORDER CHARACTERIZED BY
 CC DISPROPORTIONATE SHORT STATURE AND PLEIOTROPIC INVOLVEMENT OF THE
 CC SKELETAL AND OCULAR SYSTEMS.
 CC -!- DISEASE: DEFECTS IN COL2A1 ARE THE CAUSE OF TYPE II WAGNER
 CC SYNDROME (WS-II); A DISEASE CHARACTERIZED BY EARLY-ONSET
 CC CATARACTS, LATTICE DEGENERATION OF THE RETINA, AND RETINAL
 CC DETACHMENT WITHOUT INVOLVEMENT OF MONOCULAR TISSUES.
 CC -!- DISEASE: DEFECTS IN COL2A1 ARE THE CAUSE OF TYPE II
 CC ACHONDROGENESIS-HYPOCHONDROGENESIS; A DISEASE CHARACTERIZED BY THE
 CC ABSENCE OF OSSIFICATION IN THE VERTEBRAL COLUMN, SACRUM AND PUBIC
 CC BONES.

... Note: remainder of annotations omitted.

Query Match 67.2%; Score 45; DB 1; Length 1418;
 Best Local Similarity 66.7%; Pred. No. 4.85e+00;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 930 GPSGPGQQ 938
 ||:||||:
 QY 2 GTGEPQQE 10

RESULT 12
 ID MAM_DROME STANDARD; PRT; 1596 AA.
 AC P21519;
 DT 01-MAY-1991 (REL. 18, CREATED)
 DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
 DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
 DE NEUROGENIC PROTEIN MASTERMIND.
 GN MAM.
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CANTON-S;
 RX MEDLINE; 91065516.
 RA SMOLLER D., FRIEDEL C., SCHMID A., BETTLER D., LAM L.,
 RA YEDVOBNICK B.,
 RL GENES DEV. 4:1688-1700(1990).
 CC -!- FUNCTION: MAY HAVE A REGULATORY FUNCTION POSSIBLY IN ASSOCIATION
 CC WITH THE N GENE PRODUCT.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- DEVELOPMENTAL STAGE: DURING EARLY NEUROGENESIS MAM PRODUCTS ARE
 CC UBQUITOUSLY LOCATED. DURING LATER STAGES THEY ACCUMULATE IN THE
 CC CENTRAL NERVOUS SYSTEM.
 CC -!- THE PROTEIN HAS MANY AA HOMOPOLYMERIC DOMAINS: 21 POLY-GLN RUNS
 CC (FROM 5 TO 16 AA IN LENGTH), 4 POLY-GLY (6 TO 10 AA), 3 POLY-ASN
 CC (3 X 5 AA), 1 POLY-ALA (10 AA) AND 1 POLY-THR (5 AA) RUNS.
 CC -!- SIMILARITY: TO OTHER NUCLEAR PROTEINS OF DROSOPHILA, TO CERTAIN
 CC YEAST AND MAMMALIAN REGULATORY PROTEINS.

DR EMBL; X54251; G8204; -
 DR PIR; A33106; A33106.
 DR PIR; A36391; A36391.
 DR FLYBASE; FBgn0002643; mam.
 KW NEUROGENESIS; NUCLEAR PROTEIN; REPEAT.
 FT DOMAIN 20 84
 FT DOMAIN 127 190
 FT DOMAIN 196 219
 FT DOMAIN 259 304
 FT DOMAIN 355 388
 FT DOMAIN 392 406
 FT DOMAIN 407 440
 FT DOMAIN 651 671
 FT DOMAIN 700 714
 FT DOMAIN 759 816
 FT DOMAIN 987 996
 FT DOMAIN 1060 1079
 FT DOMAIN 1092 1107
 FT DOMAIN 1237 1252
 FT DOMAIN 1492 1496
 FT DOMAIN 1559 1592
 SQ SEQUENCE 1596 AA; 167717 MW; BEBG0500 CRC32;

Query Match 67.2%; Score 45; DB 1; Length 1596;
 Best Local Similarity 66.7%; Pred. No. 4.85e+00;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 1317 GPMGGPQQ 1325
 ||:||||:
 QY 2 GTGEPQQE 10

RESULT 13
 ID PRP2_MOUSE STANDARD; PRT; 261 AA.
 AC P05142;
 DT 13-AUG-1987 (REL. 05, CREATED)
 DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
 DE PROLINE-RICH PROTEIN MP-2 PRECURSOR.
 GN PRP.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX ANN D.K., CARLSON D.M.,
 RL J. BIOL. CHEM. 260:15863-15872(1985).
 DR EMBL; M12099; G200547; -
 DR HSSP; P19999; ICLG.
 KW REPEAT; SALIVA; SIGNAL.
 FT SIGNAL 1 15
 FT CHAIN 16 261
 FT SEQUENCE 261 AA; 26034 MW; 9D830DAF CRC32;
 SQ SEQUENCE 261 AA; 26034 MW; 9D830DAF CRC32;

Query Match 65.7%; Score 44; DB 1; Length 261;
 Best Local Similarity 75.0%; Pred. No. 8.07e+00;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 193 PTGGPQQ 200
 ||:||||:
 QY 3 PTGEPQQE 10

RESULT 14
 ID ALD_HUMAN STANDARD; PRT; 745 AA.
 AC P33897;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE ADRENOLEUKODYSTROPHY PROTEIN (ALDP).
 GN ALD.
 OS HOMO SAPIENS (HUMAN).

CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 CC EUTHERIA; PRIMATES.
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 93180910.
 RA MOSSER J., DOUAR A.-M., SARDE C.-O., KIOSCHIS P., FEIL R., MOSER H.,
 RA POUTKA A.-M., MANDEL J.-L., AUBOURG P.;
 RL NATURE 361:726-730(1993).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 9338663.
 RA REICHWALD K., SANDOVAL N., COY J., KIOSCHIS P., KORN B.,
 RA POUTKA A.-M., ROSENTHAL A.;
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 [3]
 RP REVIEW ON VARIANTS.
 RX MEDLINE: 97338663.
 RA DODD A., ROWLAND S.A., HAWKES S.L.J., KENNEDY M.A., LOVE D.R.;
 RL HUM. MUTAT. 9:500-511(1997).
 [4]
 RP VARIANT X-ALD LYS-291.
 RX MEDLINE: 94108454.
 RA CARTIER N., SARDE C.-O., DOUAR A.-M., MOSSER J., MANDEL J.-L.,
 RA AUBOURG P.;
 RL HUM. MOL. GENET. 2:1949-1951(1993).
 [5]
 RP REVIEW.
 RX MEDLINE: 93283453.
 RA AUBOURG P., MOSSER J., DOUAR A.-M., SARDE C.-O., LOPEZ J.,
 RA MANDEL J.-L.;
 RL BIOCHIMIE 75:293-302(1993).
 [6]
 RP VARIANTS X-ALD SER-148; ASP-174; ARG-266; GLN-401; TRP-418 & PHE-515.
 RX MEDLINE: 95152524.
 RA FUCHS S., SARDE C.-O., WEDEMANN H., SCHWINGER E., MANDEL J.-L.,
 RA GAL A.;
 RL HUM. MOL. GENET. 3:1903-1905(1994).
 [7]
 RP VARIANTS X-ALD TRP-518; LEU-606; CYS-617 AND HIS-617.
 RX MEDLINE: 94314951.
 RA FANEN P., GUIDOUX S., SARDE C.-O., MANDEL J.-L., GOOSSENS M.,
 RA AUBOURG P.;
 RL J. CLIN. INVEST. 94:516-520(1994).
 [8]
 RP VARIANTS X-ALD.
 RX MEDLINE: 95126139.
 RA LIGHTENBERG M.J.L., KEMP S., SARDE C.-O., VAN GEEL B.M., KLEIJER W.J.,
 RA BARTH P.G., MANDEL J.-L., VAN OOST B.A., BOLHUIS P.A.;
 RL AM. J. HUM. GENET. 56:44-50(1995).
 [9]
 RP VARIANTS X-ALD HIS-104; GLU-178; LEU-560 AND GLY-528 DEL.
 RX MEDLINE: 95233433.
 RA BRAUN A., AMBACH H., KAMMERER S., ROLINSKI B., STOECKLER S., RABL W.,
 RA GAERTNER J., ZIERZ S., ROSCHER A.A.;
 RL AM. J. HUM. GENET. 56:854-861(1995).
 [10]
 RP VARIANTS X-ALD.
 RX MEDLINE: 96047143.
 RA KOK F., NEUMANN S., SARDE C.-O., ZHENG S., WU K.-H., WEI H.-M.,
 RA BERGIN J., WATKINS P.A., GOULD S., SACK G., MOSER H., MANDEL J.-L.,
 RA SMITH K.D.;
 RL HUM. MUTAT. 6:104-115(1995).
 [11]
 RP VARIANTS X-ALD.
 RX MEDLINE: 96213748.
 RA FEIGENBAUM V., LOMBARD-PLATET G., GUIDOUX S., SARDE C.-O.,
 RA MANDEL J.-L., AUBOURG P.;
 RL AM. J. HUM. GENET. 58:1135-1144(1996).
 [12]
 RP VARIANTS X-ALD.
 RX MEDLINE: 96163493.
 RA KRAEMANN E.W., MEIER V., KORENKE G.C., HUNNEMAN D.H., HANEFELD F.;
 RL HUM. GENET. 97:194-197(1996).

[13]
 RN VARIANT AMN ARG-679.
 RP KORENKE G.C., KRAEMANN E., MEIER V., BEUCHE W., HUNNEMAN D.H.,
 RA HANEFELD F.;
 RL HUM. MUTAT. SUPPL. 1:S204-S206(1998).
 CC -I- FUNCTION: PROBABLE TRANSPORTER. COULD BE INVOLVED IN THE
 CC IMPORT OF VLCFA-COA SYNTHETASE INTO THE PEROXISOMAL MEMBRANE.
 CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL
 CC (PROBABLE).
 CC -I- DISEASE: DEFECTS IN ALD ARE THE CAUSE OF RECESSIVE X-LINKED
 CC ADRENOLEUKODYSTROPHY (X-ALD), A RARE PEROXISOMAL METABOLIC
 CC DISORDER THAT OCCURS IN BOYS AND IS CHARACTERIZED BY PROGRESSIVE
 CC MULTIFOCAL DEMYELINATION OF THE CENTRAL NERVOUS SYSTEM AND BY
 CC ADRENOCORTICAL INSUFFICIENCY. IT PRODUCES MENTAL DETERIORATION,
 CC CORTICOSPINAL TRACT DYSFUNCTION, AND CORTICAL BLINDNESS. THERE IS
 CC LABORATORY EVIDENCE OF ADRENAL CORTICAL DYSFUNCTION. DEATH
 CC INVARIABLY OCCURS IN 1 TO 5 YEARS. DIFFERENT CLINICAL
 CC MANIFESTATIONS EXIST LIKE: CEREBRAL CHILDHOOD ALD (CALD), ADULT
 CC CEREBRAL ALD (ACALD), ADRENOMYELONEUROPATHY (AMN) AND "ADDISON
 CC DISEASE ONLY" (ADO) PHENOTYPE.
 CC -I- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
 CC (ABC TRANSPORTERS). BELONGS TO THE MDR SUBFAMILY.
 CC EMBL: Z21876; G38591.
 DR EMBL: Z31348; G806557;
 DR EMBL: Z31006; G806557; JOINED.
 DR EMBL: Z31007; G806557; JOINED.
 DR EMBL: Z31008; G806557; JOINED.
 DR EMBL: Z31009; G806557; JOINED.
 DR EMBL: Z31010; G806557; JOINED.
 DR EMBL: U52111; G1302652;
 DR PIR: S30059; S30059.
 DR HSSP: P02392; ICTF.
 DR MIM: 300100;
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KW ATP-BINDING; GLYCOPROTEIN; TRANSMEMBRANE; TRANSPORT; PEROXISOME;
 KW DISEASE MUTATION; POLYMORPHISM.
 FT TRANSMEM 92 112 POTENTIAL.
 FT TRANSMEM 131 151 POTENTIAL.
 FT TRANSMEM 238 258 POTENTIAL.
 FT TRANSMEM 333 353 POTENTIAL.
 FT TRANSMEM 473 493 POTENTIAL.
 FT NP_BIND 507 514 ATP (BY SIMILARITY).
 FT VARIANT 98 98 S -> L (IN X-ALD; CALD TYPE).
 FT VARIANT 104 104 R -> C (IN X-ALD).
 FT VARIANT 104 104 R -> H (IN X-ALD; ADO-TYPE).
 FT VARIANT 105 105 T -> I (IN X-ALD; ADO-TYPE).
 FT VARIANT 107 107 L -> P (IN X-ALD; ALD/AMN/ADO-TYPES AND
 FT ASSYMPTOMATIC).
 FT VARIANT 108 108 S -> W (IN X-ALD; CALD AND AMN-TYPES).
 FT VARIANT 116 116 G -> R (IN X-ALD; CALD-TYPE).
 FT VARIANT 123 123 A -> V.
 FT VARIANT 138 141 MISSING (IN X-ALD; ALD-TYPE).
 FT VARIANT 141 141 A -> T (IN X-ALD).
 FT VARIANT 148 148 N -> S (IN X-ALD; ADO-TYPE).
 FT VARIANT 149 149 S -> N (IN X-ALD).
 FT VARIANT 152 152 R -> C (IN X-ALD; ADO-TYPE).
 FT VARIANT 152 152 R -> P (IN X-ALD).
 FT VARIANT 163 163 R -> H (IN X-ALD).
 FT VARIANT 174 174 Y -> D (IN X-ALD; ALD-TYPE).
 FT VARIANT 174 174 Y -> S (IN X-ALD; CALD-TYPE).
 FT VARIANT 178 178 Q -> E (IN X-ALD; AMN-TYPE).
 FT VARIANT 181 181 Y -> C (IN X-ALD; ALMD-TYPE).
 FT VARIANT 182 182 R -> P (IN X-ALD).
 FT VARIANT 194 194 D -> H (IN X-ALD).
 FT VARIANT 200 200 D -> V (IN X-ALD; CALD-TYPE).
 FT VARIANT 211 211 L -> P (IN X-ALD).
 FT VARIANT 220 220 L -> P (IN X-ALD).
 FT VARIANT 221 221 D -> G (IN X-ALD; CALD AND AMN-TYPES).
 FT VARIANT 254 254 T -> M (IN X-ALD; AMN-TYPE).
 FT VARIANT 254 254 T -> P (IN X-ALD; AMN-TYPE).
 FT VARIANT 263 263 P -> L (IN X-ALD; CALD, AMN AND AD-
 FT TYPES).
 FT VARIANT 266 266 G -> R (IN X-ALD).

FT VARIANT 276 276 K -> E (IN X-ALD; CALD-TYPE).
FT VARIANT 277 277 G -> R (IN X-ALD; AMN-TYPE).
FT VARIANT 277 277 G -> GN (IN X-ALD; ADO-TYPE).
FT VARIANT 277 277 G -> W (IN X-ALD).
FT VARIANT 291 291 E -> D (IN X-ALD; ACALD AND CALD-TYPES).
FT VARIANT 291 291 E -> K (IN X-ALD).
FT VARIANT 291 291 MISSING (IN X-ALD; ALD-TYPE).
FT VARIANT 291 291 A -> T (IN X-ALD; AMN-TYPE).
FT VARIANT 294 294 S -> P (IN X-ALD; AMN-TYPE).
FT VARIANT 342 342 S -> G (IN X-ALD; AMN-TYPE).
FT VARIANT 389 389 R -> H (IN X-ALD).
FT VARIANT 389 389 R -> Q (IN X-ALD; ALD AND AMN-TYPES).
FT VARIANT 401 401 R -> W (IN X-ALD; AMN-TYPE).
FT VARIANT 418 418 P -> R (IN X-ALD; CALD, AMN AND
FT VARIANT 484 484 ADO-TYPES).
FT VARIANT 507 507 G -> V (IN X-ALD; CALD-TYPES).
FT VARIANT 512 512 G -> S (IN X-ALD; CALD AND AS-TYPES).
FT VARIANT 515 515 S -> F (IN X-ALD).
FT VARIANT 518 518 R -> Q (IN X-ALD; CALD-TYPE).
FT VARIANT 518 518 R -> W (IN X-ALD; CALD-TYPE).
FT VARIANT 522 522 G -> W (IN X-ALD; AD-TYPE).
FT VARIANT 528 528 MISSING (IN X-ALD; CALD-TYPE).
FT VARIANT 534 534 P -> L (IN X-ALD; CALD-TYPE).
FT VARIANT 560 560 P -> L (IN X-ALD; CALD-TYPE).
FT VARIANT 560 560 P -> R (IN X-ALD; AMN AND ALMD-TYPES).
FT VARIANT 566 566 M -> K (IN X-ALD).
FT VARIANT 591 591 R -> Q (IN X-ALD; AMN-TYPE).
FT VARIANT 606 606 S -> L (IN X-ALD).
FT VARIANT 606 606 S -> P (IN X-ALD; CALD, AMN AND ALMD-
FT VARIANT 609 609 TYPES).
FT VARIANT 609 609 E -> G (IN X-ALD).
FT VARIANT 617 617 R -> C (IN X-ALD; AMN-TYPE).
FT VARIANT 617 617 R -> G (IN X-ALD; CALD-TYPE AND
FT VARIANT 617 617 ASYMPTOMATIC).
FT VARIANT 617 617 R -> G (IN X-ALD; ADO AND AMN-TYPES WITH
FT VARIANT 617 617 CEREBRAL INVOLVEMENT).
FT VARIANT 617 617 R -> H (IN X-ALD).
FT VARIANT 626 626 A -> T (IN X-ALD; CALD AND AMN-TYPES).
FT VARIANT 629 629 D -> H (IN X-ALD).
FT VARIANT 657 657 MISSING (IN X-ALD; CALD-TYPE).
FT VARIANT 660 660 R -> W (IN X-ALD; CALD, ALMD AND AS-
FT VARIANT 679 679 TYPES).
FT VARIANT 679 679 W -> R (IN X-ALD; AMN-TYPE).
FT SEQUENCE 745 AA; 82908 MW; F37A45D3 CRC32;

Query Match 65.7%; Score 44; DB 1; Length 745;
Best Local Similarity 60.0%; Pred. No. 8.07e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 47 QAPAGEPTQE 56
QY 1 DGPTGEPTQE 10

RESULT 15
ID RRE1_HUMAN STANDARD; PRT; 755 AA.
AC Q92766;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE RAS-RESPONSIVE ELEMENT BINDING PROTEIN 1 (RREB-1).
GN RREB1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-THYROID CARCINOMA;
RX MEDLINE; 96413283.
RA THIAGALINGAM A., BAYLIN S.B., BORGES M., NELKIN B.D.;
RL MOL. CELL. BIOL. 16:5335-5345(1996).
CC -!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS SPECIFICALLY TO THE
DISTAL RAS-RESPONSIVE ELEMENT (RRE) IN THE CALCITONIN GENE

CC PROMOTER AND AUGMENT THE RAS/RAF-MEDIATED TRANSCRIPTIONAL RESPONSE
CC OF THAT PROMOTER. MAY BE INVOLVED IN RAS/RAF-MEDIATED CELL
CC DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
DR EMBL; U26914; GI654112; -.
DR MIM; 602209; -.
DR PROSITE; PS00028; ZINC_FINGER_C2H2; 4.
KW TRANSCRIPTION REGULATION; ACTIVATOR; ZINC-FINGER; METAL-BINDING;
KW DNA-BINDING; NUCLEAR PROTEIN.
FT ZN_FING 315 337 C2H2-TYPE.
FT ZN_FING 461 483 C2H2-TYPE.
FT ZN_FING 580 602 C2H2-TYPE.
FT ZN_FING 608 630 C2H2-TYPE.
SQ SEQUENCE 755 AA; 79865 MW; 50627236 CRC32;

Query Match 65.7%; Score 44; DB 1; Length 755;
Best Local Similarity 50.0%; Pred. No. 8.07e+00;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 501 EGSPAPEQEQ 510
QY 1 DGPTGEPTQE 10

Search completed: Thu May 13 15:37:02 1999
Job time : 8 secs.


```
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NAROO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RL DNA RES. 3:109-136(1996).
DR EMBL: D90910; G1652997; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 414 AA; 44810 MW; 0E3A0E9E CRC32;

Query Match 71.6%; Score 48; DB 2; Length 414;
Best Local Similarity 60.0%; Pred. No. 1.74e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 318 EGPTEAQQQ 327
QY 1 DGPTGEPOQE 10

RESULT 3
ID P79765 PRELIMINARY; PRT; 357 AA.
AC P79765;
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE NEUROD-LIKE PROTEIN.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC GALLIFORMES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEHORN; TISSUE-RETINA;
RA ROZTOCIL T., MATTER-SADZINSKI L., ALLIOD C., BALLIVET M., MATTER J.M.;
RL DEVELOPMENT 124:3263-3272(1997).
DR EMBL: Y09596; E283389; -.
DR PFAM: PF00010; HLH.
SQ SEQUENCE 357 AA; 38809 MW; 04A96823 CRC32;

Query Match 70.1%; Score 47; DB 13; Length 357;
Best Local Similarity 66.7%; Pred. No. 2.95e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 9 GPAGEPQAQ 17
QY 2 GPTGEPOQE 10

RESULT 4
ID 073826 PRELIMINARY; PRT; 357 AA.
AC 073826;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE NEUROD.
GN NEUROD.
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC GALLIFORMES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WHITE LEHORN;
RA YAN R.-T., WANG S.-Z.;
RL J. NEUROBIOL. 0:0-0(1998).
DR EMBL: AF060885; G3094020; -.
SQ SEQUENCE 357 AA; 38858 MW; 7D406FF3 CRC32;

Query Match 70.1%; Score 47; DB 13; Length 357;
Best Local Similarity 66.7%; Pred. No. 2.95e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 9 GPAGEPQAQ 17
```

```
QY 2 GPTGEPOQE 10

RESULT 5
ID 044431 PRELIMINARY; PRT; 676 AA.
AC 044431;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE CIRCADIAN CLOCK PROTEIN (FRAGMENT).
GN TIM.
OS DROSOPHILA HYDEI (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN [1]
RP SEQUENCE FROM N.A.
RA OUSLEY A., ZAFARULLAH K., CHEN Y., EMERSON M., HICKMAN L., SEHGAL A.;
RL GENETICS 0:0-0(1998).
DR EMBL: AF038579; G2734130; -.
FT NON_TER 1
FT NON_TER 676
SQ SEQUENCE 676 AA; 76022 MW; 83E96D47 CRC32;

Query Match 70.1%; Score 47; DB 5; Length 676;
Best Local Similarity 60.0%; Pred. No. 2.95e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 357 DGPGKQHQ 366
QY 1 DGPTGEPOQE 10

RESULT 6
ID Q20778 PRELIMINARY; PRT; 1120 AA.
AC Q20778;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE SIMILAR TO TRIPLE HELICAL REGION OF COLLAGENS.
GN F54D8.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACLOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RL NATURE 0:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA BENTLEY D.;
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U12966; G529221; -.
SQ SEQUENCE 1120 AA; 119368 MW; E0352B8E CRC32;

Query Match 70.1%; Score 47; DB 5; Length 1120;
Best Local Similarity 60.0%; Pred. No. 2.95e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

Db 1020 EGTGPGCAD 1029
 QY :||||| :
 1 DGPTGEPQOE 10

RESULT 7
 ID O4430 PRELIMINARY; PRT; 1194 AA.
 AC O4430;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE CIRCADIAN CLOCK PROTEIN (FRAGMENT).
 GN TIM.
 OS DROSOPHILA VIRILIS (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA OUSLEY A., ZAFARULLAH K., CHEN Y., EMERSON M., HICKMAN L., SERGAL A.;
 RL GENETICS 0:0-0(1998).
 DR EMBL; AF038502; G2746733; -.
 FT NON_TER 1194 1194
 SQ SEQUENCE 1194 AA; 134946 MW; F6C4345B CRC32;

Query Match 70.18; Score 47; DB 5; Length 1194;
 Best Local Similarity 60.0%; Pred. No. 2.95e+00;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Db 486 DGPGKPOHQ 495
 QY :||||| :
 1 DGPTGEPQOE 10

RESULT 8
 ID O17482 PRELIMINARY; PRT; 1343 AA.
 AC O17482;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE TIMELESS (TIM).
 GN TIM.
 OS DROSOPHILA VIRILIS (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 98033379.
 RA MYERS M.P., ROTHENFLUH A., CHANG M., YOUNG M.W.;
 RL NUCLEIC ACIDS RES. 25:4710-4714(1997).
 DR EMBL; AF032403; G2641617; -.
 DR EMBL; AF032402; G2641617; JOINED.
 SQ SEQUENCE 1343 AA; 150942 MW; 77573CAC CRC32;

Query Match 70.18; Score 47; DB 5; Length 1343;
 Best Local Similarity 60.0%; Pred. No. 2.95e+00;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Db 486 DGPGKPOHQ 495
 QY :||||| :
 1 DGPTGEPQOE 10

RESULT 9
 ID O44380 PRELIMINARY; PRT; 1398 AA.
 AC O44380;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE CIRCADIAN CLOCK PROTEIN.
 GN TIM.
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 98033379.

RA MYERS M.P., ROTHENFLUH A., CHANG M., YOUNG M.W.;
 RL NUCLEIC ACIDS RES. 25:4710-4714(1997).
 DR EMBL; AF032401; G2655282; -.
 DR EMBL; AF032400; G2655282; JOINED.
 SQ SEQUENCE 1398 AA; 156366 MW; 59A88D49 CRC32;

Query Match 70.1%; Score 47; DB 5; Length 1398;
 Best Local Similarity 60.0%; Pred. No. 2.95e+00;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Db 519 DGPGKPOHQ 528
 QY :||||| :
 1 DGPTGEPQOE 10

RESULT 10
 ID O35966 PRELIMINARY; PRT; 231 AA.
 AC O35966;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE SEX DETERMINING PROTEIN.
 GN SRI.
 OS MUS MUSCULUS DOMESTICUS (WESTERN EUROPEAN HOUSE MOUSE).
 OC EUKARYOTA; MITOCHONDRIAL EUKARYOTES; METAZOA; VERTEBRATA;
 OC MAMMALIA; EUTHERIA; RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS;
 OC MUS MUSCULUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-B6JEI-YBUB, CD-1 AND B6JEI-YKAM;
 RX MEDLINE; 98043417.
 RA ALBRECHT K.H., EICHER E.M.;
 RL GENETICS 147:1267-1277(1997).
 DR EMBL; AF009519; G2271479; -.
 DR EMBL; U70643; G2623351; -.
 DR EMBL; U70644; G2623353; -.
 DR PFAM; PF00505; HMG_box.
 SQ SEQUENCE 231 AA; 28533 MW; 66F6EDC1 CRC32;

Query Match 68.7%; Score 46; DB 11; Length 231;
 Best Local Similarity 70.0%; Pred. No. 4.96e+00;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Db 137 DIPTGHPQOQ 146
 QY :||||| :
 1 DGPTGEPQOE 10

RESULT 11
 ID O35968 PRELIMINARY; PRT; 232 AA.
 AC O35968;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE SEX DETERMINING PROTEIN.
 GN SRI.
 OS MUS MUSCULUS DOMESTICUS (WESTERN EUROPEAN HOUSE MOUSE).
 OC EUKARYOTA; MITOCHONDRIAL EUKARYOTES; METAZOA; VERTEBRATA;
 OC MAMMALIA; EUTHERIA; RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS;
 OC MUS MUSCULUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-B6JEI-YWSB, B6JEI-YAPP AND B6JEI-YAKR;
 RX MEDLINE; 98043417.
 RA ALBRECHT K.H., EICHER E.M.;
 RL GENETICS 147:1267-1277(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-B6-YAKR; SPECIES=WESTERN EUROPEAN HOUSE MOUSE;
 RX MEDLINE; 96207297.
 RA CARLSLE C., WINKING H., WEICHENHAN D., NAGAMINE C.M.;
 RL GENOMICS 33:32-45(1996).
 RN [3]

RP SEQUENCE FROM N.A.
RC STRAIN=B6-YARR; SPECIES=WESTERN EUROPEAN HOUSE MOUSE;
RA CARLISLE C., NAGAMINE C.M.;
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U70650; G2623365; -;
DR EMBL; U70641; G2623347; -;
DR EMBL; U70642; G2623349; -;
DR EMBL; AF068053; G3176975; -;
DR PFAM; PF00505; HMG_box
SQ SEQUENCE 232 AA; 28790 MW; 6998BFF4 CRC32;

Query Match 68.7%; Score 46; DB 11; Length 232;
Best Local Similarity 70.0%; Pred. No. 4.96e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 137 DIPTGHPQQ 146
| ||| |||
QY 1 DGPTGEPOQE 10

RESULT 12 PRELIMINARY; PRT; 233 AA.
ID O35860
AC O35860;

DT 01-JAN-1998 (TREMREL. 05, CREATED)
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)
DE SEX DETERMINING PROTEIN.
GN SRY.

OS MUS MUSCULUS DOMESTICUS (WESTERN EUROPEAN HOUSE MOUSE).
OC EUKARYOTAE; MITOCHONDRIAL EUKARYOTES; METAZOA; CHORDATA; VERTEBRATA;
OC MAMMALIA; EUTHERIA; RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS;
OC MUS MUSCULUS.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=WMP/EI;
RX MEDLINE; 98043417.
RA ALBRECHT K.H., EICHER E.M.;
RL GENETICS 147:1267-1277(1997).
DR EMBL; U70649; G2623363; -;
DR PFAM; PF00505; HMG_box.
SQ SEQUENCE 233 AA; 28790 MW; 3F2451C1 CRC32;

Query Match 68.7%; Score 46; DB 11; Length 233;
Best Local Similarity 70.0%; Pred. No. 4.96e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 137 DIPTGHPQQ 146
| ||| |||
QY 1 DGPTGEPOQE 10

RESULT 13 PRELIMINARY; PRT; 234 AA.
ID O35858
AC O35858;

DT 01-JAN-1998 (TREMREL. 05, CREATED)
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)
DE SEX DETERMINING PROTEIN.
GN SRY.

OS MUS MUSCULUS DOMESTICUS (WESTERN EUROPEAN HOUSE MOUSE).
OC EUKARYOTAE; MITOCHONDRIAL EUKARYOTES; METAZOA; CHORDATA; VERTEBRATA;
OC MAMMALIA; EUTHERIA; RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS;
OC MUS MUSCULUS.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=B6JEI-YLIP;
RX MEDLINE; 98043417.
RA ALBRECHT K.H., EICHER E.M.;
RL GENETICS 147:1267-1277(1997).
DR EMBL; U70645; G2623355; -;
DR PFAM; PF00505; HMG_box.
SQ SEQUENCE 234 AA; 28918 MW; 2446EF1B CRC32;

Query Match 68.7%; Score 46; DB 11; Length 234;
Best Local Similarity 70.0%; Pred. No. 4.96e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 137 DIPTGHPQQ 146
| ||| |||
QY 1 DGPTGEPOQE 10

RESULT 14 PRELIMINARY; PRT; 234 AA.
ID O35859
AC O35859;

DT 01-JAN-1998 (TREMREL. 05, CREATED)
DT 01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)
DE SEX DETERMINING PROTEIN.
GN SRY.

OS MUS MUSCULUS DOMESTICUS (WESTERN EUROPEAN HOUSE MOUSE).
OC EUKARYOTAE; MITOCHONDRIAL EUKARYOTES; METAZOA; CHORDATA; VERTEBRATA;
OC MAMMALIA; EUTHERIA; RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS;
OC MUS MUSCULUS.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=B6JEI-YORB;
RX MEDLINE; 98043417.
RA ALBRECHT K.H., EICHER E.M.;
RL GENETICS 147:1267-1277(1997).
DR EMBL; U70646; G2623357; -;
DR PFAM; PF00505; HMG_box.
SQ SEQUENCE 234 AA; 28918 MW; 69C12948 CRC32;

Query Match 68.7%; Score 46; DB 11; Length 234;
Best Local Similarity 70.0%; Pred. No. 4.96e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 137 DIPTGHPQQ 146
| ||| |||
QY 1 DGPTGEPOQE 10

RESULT 15 PRELIMINARY; PRT; 238 AA.
ID O69004
AC O69004;

DT 01-AUG-1998 (TREMREL. 07, CREATED)
DT 01-AUG-1998 (TREMREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
DE FERRISIDROPHORE RECEPTOR-LIKE PROTEIN.
GN PIGC.

OS PSEUDOMONAS AERUGINOSA.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC PSEUDOMONADACEAE.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=PAOI;
RX MEDLINE; 96210657.
RA OCHSNER U.A., VASIL M.L.;
RL PROC. NATL. ACAD. SCI. U.S.A. 93:4409-4414(1996).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN=PAOI;
RA OCHSNER U.A., JOHNSON Z., VASIL A.I., VASIL M.L.;
RL SUBMITTED (APR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF060193; G3091148; -;
SQ SEQUENCE 238 AA; 25049 MW; 582EDF82 CRC32;

Query Match 68.7%; Score 46; DB 2; Length 238;
Best Local Similarity 55.6%; Pred. No. 4.96e+00;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 182 GSTGDPQRD 190
| : : : : :
QY 2 GPTGEPOQE 10

Search completed: Thu May 13 15:37:55 1999
Job time : 34 secs.

M P E R E L

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 13 15:39:59 1999; MasPar time 2.44 Seconds

Tabular output not generated. 39.713 Million cell updates/sec

Title: >US-09-040-485-8
Description: (3-8) from US09040485.pep (2 of 2)
Perfect Score: 41
Sequence: 1 PTGEPQ 6

Scoring table: PAM 150
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29

Statistics: Mean 12.609; Variance 30.187; scale 0.418

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	38	92.7	539 14	R80506	S. lividans protease	8.16e+01
2	37	90.2	3038 14	R74171	Aspergillus terreus t	1.11e+02
3	35	85.4	512 3	R14908	ORF from BIK1-FUS1 re	2.06e+02
4	35	85.4	512 3	P94880	FUS 1 promoter incorp	2.06e+02
5	35	85.4	512 3	R14909	FUS1 protein.	2.06e+02
6	35	85.4	1931 5	R27649	Human calcium channel	2.06e+02
7	35	85.4	2237 6	R33550	Sequence of the alpha	2.06e+02
8	35	85.4	2237 14	R71006	Human neuronal calciu	2.06e+02
9	35	85.4	2339 6	R33549	Sequence of the alpha	2.06e+02
10	35	85.4	2339 14	R71005	Human neuronal calciu	2.06e+02
11	34	82.9	59 25	W32784	Rat thymus 60K-CSF pr	2.80e+02
12	34	82.9	109 20	W02134	High mobility group H	2.80e+02
13	34	82.9	177 16	R81430	Hepatitis GB virus (H	2.80e+02
14	34	82.9	343 20	W02132	High mobility group p	2.80e+02
15	34	82.9	649 29	W38218	Maize starch synthase	2.80e+02
16	34	82.9	659 20	W06708	Mouse haematopoietic-	2.80e+02
17	34	82.9	680 2	R07433	First open reading fr	2.80e+02
18	34	82.9	757 20	W03179	Bovine poly-immunoglo	2.80e+02

19	34	82.9	806 5	R27481	RP-III residual prote	2.80e+02
20	34	82.9	1403 10	R54629	Mannuronan C-5-epimer	2.80e+02
21	34	82.9	1838 10	R53257	Human collagen (type	2.80e+02
22	34	82.9	2322 22	W15566	Melanoma-associated c	2.80e+02
23	34	82.9	3567 8	R44431	eryA region polypepti	2.80e+02
24	34	80.5	46 24	W16397	HSV-1 glycoprotein B	3.78e+02
25	33	80.5	167 4	R22273	Human collagen XI ins	3.78e+02
26	33	80.5	400 2	P70654	Sequence encoded by g	3.78e+02
27	33	80.5	442 20	W06709	Human haematopoietic-	3.78e+02
28	33	80.5	454 4	R25155	Fibronectin-spacer-co	3.78e+02
29	33	80.5	478 9	R47586	Human bone formation-	3.78e+02
30	33	80.5	478 20	W06539	Human bone morphogen	3.78e+02
31	33	80.5	579 20	W08143	RHDV capsid protein.	3.78e+02
32	33	80.5	659 17	R94534	BTK tyrosine kinase.	3.78e+02
33	33	80.5	903 2	P70426	Recombinant herpes si	3.78e+02
34	33	80.5	903 3	P60244	Herpes simplex virus	3.78e+02
35	33	80.5	903 3	P71135	Herpes Simplex Virus-	3.78e+02
36	33	80.5	904 8	R41779	Glycoprotein B (gB2).	3.78e+02
37	33	80.5	904 20	W00375	HSV-1 glycoprotein B.	3.78e+02
38	33	80.5	904 26	W34553	Herpes simplex virus	3.78e+02
39	33	80.5	905 3	P80915	Sequence of Herpes si	3.78e+02
40	33	80.5	1058 9	R42453	Enzyme involved in ei	3.78e+02
41	33	80.5	1366 13	R17102	Collagen alpha 2 (I)	3.78e+02
42	33	80.5	2756 20	R99462	Biosynthetic enzyme o	3.78e+02
43	33	80.5	3080 1	P93285	Sequence of clone HIV	3.78e+02
44	33	80.5	3210 1	P81771	Deduced sequence enco	3.78e+02
45	32	78.0	497 17	R81462	Human derived cytochr	5.09e+02

ALIGNMENTS

RESULT 1
ID R80506 standard; Protein; 539 AA.
AC R80506;
DT 04-DEC-1995 (first entry)
DE S. lividans protease p5-6.
KW Protease; metalloendoproteinase; tripeptidyl aminopeptidase;
OS protease-deficiency; protein secretion.
FS Streptomyces lividans.
FH Key Location/Qualifiers
FT peptide 1..47
FT /label= Sig_peptide
PN W09517512-A.
PD 29-JUN-1995.
PF 22-DEC-1994; U14772.
PR 23-DEC-1993; US-173508.
PA (CANG-) CANGENE CORP.
PI Bartfeld D, Butler MJ, Hadary D, Jenish DL, Krieger TJ;
PI Malek LT, Soostmeyer G, Walczyk E;
DR WPI; 95-240673/31.
DR N-PSDB; Q99366.
PT Endogenous Streptomyces protease(s), opt. having impaired activity -
PT useful in prodn. of exogenous proteins with reduced proteolytic
PT degradation
PS Claim 4; Fig.14; 142pp; English.
CC Protease Tap-negative cells were transformed with a S. lividans 66
CC genomic library and screened with APA-beta-naphthylamide to
CC isolate colonies contg. genes (Q99365-68) for novel proteases
CC P5-4, P5-6, P5-10 and P8-2 (R80505-08). Impaired expression of
CC such proteases by Streptomyces hosts improves the quality, quantity
CC and stability of exogenous gene products.
SQ Sequence 539 AA;

Query Match 92.7%; Score 38; DB 14; Length 539;
Best Local Similarity 83.3%; Pred No. 8.16e+01;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 457 ptgeph 462

QY 3 PTGEPQ 8

RESULT 2

ID R74171 standard; Protein; 3038 AA.
AC R74171;
DT 18-JAN-1996 (first entry)
DE Aspergillus terreus triol polyketide synthase.
KW triol polyketide synthase; TPKS; HMG-CoA reductase inhibitor;
KW hypercholesterolaemia; LDL-cholesterol.
OS Aspergillus terreus.
FH Key
FT Location/Qualifiers
FT region 181
FT /label= keto-acyl synthase motif
FT region 654..658
FT /label= acetyl/malonyl transferase motif
FT region 985..994
FT /label= dehydratase motif
FT region 1446..1450
FT /label= methyl transferase motif
FT region 1932..1937
FT /label= enoyl reductase motif
FT region 2164..2169
FT /label= keto reductase motif
FT region 2498
FT /label= acyl carrier protein motif
FT misc_difference 282..288
FT /label= misc feature
FT misc_difference 1450..1460
FT /label= misc feature
FT misc_difference 1603..1612
FT /label= misc feature
FT misc_difference 2521..2535
FT /label= misc feature
FT WO9512661-A.
PN 11-MAY-1995.
PD 28-OCT-1994; U12423.
PF 02-NOV-1993; US-148132.
PR (MERI) MERCK & CO INC.
PA Conder MJ, Davis CR, Hendrickson LE, Mcada PC, Rambosek J;
PI Reeves CD, Vinci VA;
DR WPI: 95-193816/25.
DT N-PSDB; Q92323.
PR Novel DNA encoding triol polyketide synthase - used to isolate and
PT identify homologues of triol polyketide synthase, and in the treatment
PT of hyper-cholesterolaemia
PS Claim 12; Figure 2; 107pp; English.
CC The full-length TPKS-encoding DNA in plasmid pLOA was
CC designated pPKS100. Splicing of the introns from the DNA
CC sequence and translation of the 914 nt ORF results in a
CC protein of 3038 AAs (R74171) with a mol. wt. of 269,090
CC daltons. Inspection of the TPKS AA sequence for active
CC site residues and motifs known to be associated with
CC polyketide synthases and fatty acid synthase (FAS) activities
CC resulted in the identification of candidates for expected
CC sites (see FT). Except for the presence of a methyl transferase,
CC not present in FAS, the succession of activities on the
CC TPKS protein is the same as that observed for the rat FAS
CC protein.
CC Sequence 3038 AA;
SQ
Query Match 90.2%; Score 37; DB 14; Length 3038;
Best Local Similarity 66.7%; Pred. No. 1.11e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 325 pagdpq 330
|:|:|
QY 3 PTGEPQ 8
RESULT 3
ID R14908 standard; Protein; 512 AA.
AC R14908;
DT 03-FEB-1992 (first entry)
DE ORF from BIK1-FUS1 region.
KW Pheromone inducible yeast promoter; bilateral karyogamy defect;
KW FUS2; BIK1.

OS Saccharomyces cerevisiae.
PN US5063154-A.
PD 05-NOV-1991.
PF 24-JUN-1988; 212270.
PR 24-JUN-1987; US-066078.
PR 24-JUN-1988; US-212270.
PA (WHIT-) WHITEHEAD INST BIOM.
PI Fink GR, Trueheart J, Elion EA;
DR WPI: 91-346534/47.
DR N-PSDB; Q14754.
DT DNA fragment contg. pheromone-inducible yeast promoter - useful
PT for transforming yeast cells to produce foreign proteins, which
PT may be toxic to yeast cells.
PS Disclosure: Fig 4; 23pp; English.
CC The ORF occurs between the BIK1 reading frame (on the complementary
CC strand) and the FUS1 reading frame on chromosome III of S. cerevis-
CC iae. Transcription of the FUS1 gene is greatly enhanced by the
CC presence of the appropriate mating pheromone. The promoter region
CC can be used for the pheromone inducible expression of proteins
CC of interest.
CC See also R14907-10.
CC Sequence 512 AA;
SQ
Query Match 85.4%; Score 35; DB 3; Length 512;
Best Local Similarity 83.3%; Pred. No. 2.06e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 282 ptgapq 287
|:|:|
QY 3 PTGEPQ 8
RESULT 4
ID P94880 standard; Protein; 512 AA.
AC P94880;
DT 02-JUL-1990 (first entry)
DE FUS 1 promoter incorporated within the HIS4 gene.
DE Yeast promoter; yeast pheromones; FUS-1; BIK-1; HIS-4.
OS Saccharomyces cerevisiae.
PN WO8810308-A.
PD 29-DEC-1988.
PF 23-JUN-1988; 02129.
PR 24-JUN-1987; US-066078.
PA (WHIT-) Whitehead Inst.
PI Fink GR, Trueheart J, Elion EA;
DR WPI: 89-023850/03.
DR N-PSDB: N93100.
DT New DNA fragment contg. protein encoding gene and yeast promoter -
PT controlled by mating pheromone allowing efficient and
PT regulatable expression.
PS Claim 5; Fig 4; 51pp; English.
CC FUS 1 promoter found within HIS4 gene may be used to promote at least one
CC polypeptide within a high-copy vector induced by a-factor for alpha cells
CC and alpha-factor for a-cells.
CC Sequence 512 AA;
SQ
Query Match 85.4%; Score 35; DB 1; Length 512;
Best Local Similarity 83.3%; Pred. No. 2.06e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 282 ptgapq 287
|:|:|
QY 3 PTGEPQ 8
RESULT 5
ID R14909 standard; Protein; 512 AA.
AC R14909;
DT 03-FEB-1992 (first entry)
DE FUS1 protein.
KW Pheromone inducible yeast promoter; bilateral karyogamy defect;
KW FUS2; BIK1.
OS Saccharomyces cerevisiae.

PN US063154-A.
 PD 05-NOV-1991.
 PF 24-JUN-1988; 212270.
 PR 24-JUN-1987; US-066078.
 PR 24-JUN-1988; US-212270.
 PA (WHT-) WHITEHEAD INST BIOM.
 PI Fink GR, Trueheart J, Ellison EA;
 DR WPI; 91-346534/47.
 DR N-PSDB; Q14754.
 PT DNA fragment contg. pheromone-inducible yeast promoter - useful
 PT for transforming yeast cells to produce foreign proteins, which
 PT may be toxic to yeast cells.
 PS Disclosure; Fig 4; 23pp; English.
 CC The FUS1 gene is involved in the initial zygote formation (cell
 CC surface recognition leading to cytoplasmic fusion) during conju-
 CC gation. Transcription of the FUS1 gene is greatly enhanced by the
 CC presence of the appropriate mating pheromone. The promoter region
 CC can be used for the pheromone inducible expression of proteins
 CC of interest.
 CC See also RI4907-10.
 CC Sequence 512 AA;
 SQ

Query Match 85.4%; Score 35; DB 3; Length 512;
 Best Local Similarity 83.3%; Pred. No. 2.06e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 282 ptgapq 287
 |||||
 QY 3 PTGEPQ 8

RESULT 6
 ID R27649 standard; Protein; 1931 AA.
 AC R27649.
 DT 03-MAR-1993 (first entry)
 DE Human calcium channel 27980/11.
 KW Plasmid pRI4-5.3.3.1; Ca-flux assay.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_difference 1165
 FT /note= "encoded by GNG codon, N is unknown"
 PN EP-507170-A.
 PD 07-OCT-1992.
 PF 23-MAR-1992; 104970.
 PR 04-APR-1991; DE-110785.
 PA (FARB) BAYER AG.
 PI Franz J, Rae P, Unterbeck A, Weingaertner B;
 DR WPI; 92-333446/41.
 DR N-PSDB; Q29269.
 PT Cloned human neuronal calcium channel sub-types - useful in
 PT calcium flux assays to screen for neurone-specific calcium
 PT channel ligands
 PS Claim 2; Page 63-77; 101pp; German.
 CC Human neuroblastoma cell line, hippocampus, frontal and temporal
 CC cortex and visual cortex cDNA banks were screened with a probe
 CC containing carp skeletal muscle Ca-channel cDNA. The cDNA clone
 CC pRI4-5.3.3.1 overlaps with clone p1247-14.1.1.1 (see Q29263).
 CC There were a number of differences between the two sequences
 CC including the deletion of an Adenosine residue at position 1013 of
 CC p1247-14.1.1.1 which leads to a stop codon at position 1028-1030;
 CC the deletion is thought to be a cloning artefact. The human
 CC neuronal calcium channel protein can be used for screening for Ca
 CC channel ligands (agonists or antagonists). See also Q29259-Q29275.
 SQ Sequence 1931 AA;

Query Match 85.4%; Score 35; DB 5; Length 1931;
 Best Local Similarity 83.3%; Pred. No. 2.06e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1577 pdgepq 1582
 |||||
 QY 3 PTGEPQ 8

RESULT 7
 ID R33550 standard; Protein; 2237 AA.
 AC R33550;
 DT 30-JUN-1993 (first entry)
 DE Sequence of the alpha 1B-2 human calcium channel subunit.
 KW Human calcium channel subunit; diagnosis; agonist; antagonist;
 KW Lambert Eaton syndrome.
 OS Homo sapiens.
 PN W09304083-A.
 PD 04-MAR-1993.
 PF 14-AUG-1992; U06903.
 PR 15-AUG-1991; US-745206.
 PR 10-APR-1992; US-868354.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, Mccue AF,
 PI Williams ME;
 DR WPI; 93-093936/11.
 DR N-PSDB; Q37818.
 PT DNA encoding specific human calcium channel sub-units - used for
 PT identifying calcium channel agonists and antagonists and
 PT diagnosing Lambert Eaton syndrome
 PS Disclosure; Page 120-128; 150pp; English.
 CC DNA encoding the alpha 1B subunit was isolated by screening a
 CC human basal ganglia cDNA library with fragments of the rabbit
 CC skeletal muscle calcium channel alpha 1 subunit-encoding cDNA.
 CC A portion of one of the positive clones was used to screen an IMR32
 CC cell cDNA library. Clones that hybridized to the basal ganglia
 CC DNA prove were used to further screen an IMR32 cell cDNA library
 CC to identify overlapping clones that in turn were used to screen a
 CC human hippocampus cDNA library. In this way, a sufficient series of
 CC clones to span nearly the entire length of the nucleotide sequence
 CC encoding the human alpha 1B subunit was obtained. PCR amplification
 CC of specific regions of the IMR32 cell alpha 1B mRNA yielded
 CC additional segments of the alpha 1B coding sequence. A full-length
 CC alpha 1B DNA clone was constructed by ligating portions of the
 CC partial cDNA clones (see Q37817, Q37818). Alpha 1B-1 and alpha
 CC 1B-2 are derived by alternative splicing of the alpha 1B subunit
 CC transcript.
 SQ Sequence 2237 AA;

Query Match 85.4%; Score 35; DB 6; Length 2237;
 Best Local Similarity 83.3%; Pred. No. 2.06e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1985 pdgepq 1990
 |||||
 QY 3 PTGEPQ 8

RESULT 8
 ID R71006 standard; Protein; 2237 AA.
 AC R71006;
 DT 01-DEC-1995 (first entry)
 DE Human neuronal calcium channel subunit alpha 1B-2.
 KW Calcium channel subunit; antagonist; agonist; diagnosis;
 KW Lambert Eaton Syndrome.
 OS Homo sapiens.
 PN W09504822-A.
 PD 16-FEB-1995.
 PF 11-AUG-1994; U09230.
 PR 11-AUG-1993; US-105536.
 PR 05-NOV-1993; US-149097.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Ellis SB, Gillespie A, Harpold MM, Mccue AF, Williams ME;
 DR WPI; 95-090900/12.
 DR N-PSDB; Q84658.
 PT DNA encoding human calcium channel sub-unit(s) - used for
 PT developing prods. for studying calcium channels, e.g. for
 PT obtaining agonists and antagonists
 PS Disclosure; Page 149-160; 285pp; English.
 CC DNA encoding the alpha 1B subunit was isolated by screening a
 CC human basal ganglia cDNA library with fragments of the rabbit

CC skeletal muscle calcium channel alpha 1 subunit-encoding cDNA.
 CC A portion of one of the positive clones was used to screen an
 CC IMR32 cell cDNA library. Clones that hybridized to the basal
 CC ganglia probe were used to further screen an IMR32 cell cDNA
 CC library to identify overlapping clones that in turn were used
 CC to screen a human hippocampus cDNA library. A series of clones
 CC to span nearly the entire length of the nt. sequence encoding
 CC the human alpha 1B subunit was obtained. Nucleic acid amplification
 CC of specific regions of the IMR32 cell alpha 1B mRNA yielded
 CC additional segments of the alpha 1B coding sequence. A full-
 CC length alpha 1B DNA clone was constructed by ligating portions
 CC of the partial cDNA clones. Nucleic acid amplification analysis
 CC of IMR32 cell RNA and genomic DNA using oligo primers correspond
 CC to sequences located 5' and 3' of the stop codon of the DNA encoding
 CC the alpha 1B subunit revealed an alternatively spliced alpha
 CC 1B-encoding mRNA in IMR32 cells. This second mRNA product is the
 CC result of differential splicing of the alpha 1B subunit transcript
 CC to include another exon that is not present in the mRNA correspond
 CC to the other 3' alpha 1B cDNA sequence that was initially isolated.
 CC The alpha 1B subunit encoded by a DNA sequence contig. an additional
 CC exon is referred to as alpha 1B-1 and given in Q84657/R71005,
 CC whereas the other form is referred to as alpha 1B-2 and is given in
 CC Q84658/R71006. Following the sequence of the additional exon in
 CC alpha 1B-1 the alpha 1B-1 and alpha 1B-2 sequences are identical.
 CC Sequence 2237 AA;

Query Match 85.4%; Score 35; DB 14; Length 2237;
 Best Local Similarity 83.3%; Pred. No. 2.06e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1985 pdgepq 1990
 | ||||
 Qy 3 PTGEPQ 8

RESULT 9
 ID R33549 standard; Protein; 2339 AA.
 AC R33549;
 DE 30-JUN-1993 (first entry)
 DE Sequence of the alpha 1B-1 human calcium channel subunit.
 KW Human calcium channel subunit; diagnosis; agonist; antagonist;
 KW Lambert Eaton syndrome.
 OS Homo sapiens.
 PN WQ9304083-A.
 PD 04-MAR-1993.
 PF 14-AUG-1992; U06903.
 PR 15-AUG-1991; US-745206.
 PR 10-APR-1992; US-868354.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Brenner R, Ellis SB, Feldman DH, Harpold MM, McCue AF,
 PI Williams ME;
 DR WPI; 93-093936/11.
 DR N-PSDB; Q37817.
 PT DNA encoding specific human calcium channel sub-units - used for
 PT identifying calcium channel agonists and antagonists and
 PT diagnosing Lambert Eaton syndrome
 PS Disclosure; Page 111-120; 150pp; English.
 CC DNA encoding the alpha 1B subunit was isolated by screening a
 CC human basal ganglia cDNA library with fragments of the rabbit
 CC skeletal muscle calcium channel alpha 1 subunit-encoding cDNA.
 CC A portion of one of the positive clones was used to screen an IMR32
 CC cell cDNA library. Clones that hybridized to the basal ganglia
 CC DNA probe were used to further screen an IMR32 cell cDNA library
 CC to identify overlapping clones that in turn were used to screen a
 CC human hippocampus cDNA library. In this way, a sufficient series of
 CC clones to span nearly the entire length of the nucleotide sequence
 CC encoding the human alpha 1B subunit was obtained. PCR amplification
 CC of specific regions of the IMR32 cell alpha 1B mRNA yielded
 CC additional segments of the alpha 1B coding sequence. A full-length
 CC alpha 1B DNA clone was constructed by ligating portions of the
 CC partial cDNA clones (see Q37817, Q37818). Alpha 1B-1 and alpha
 CC 1B-2 are derived by alternative splicing of the alpha 1B subunit
 CC transcript.

SQ Sequence 2339 AA;
 Query Match 85.4%; Score 35; DB 6; Length 2339;
 Best Local Similarity 83.3%; Pred. No. 2.06e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1985 pdgepq 1990
 | ||||
 Qy 3 PTGEPQ 8

RESULT 10
 ID R71005 standard; Protein; 2339 AA.
 AC R71005;
 DE 01-DEC-1995 (first entry)
 DE Human neuronal calcium channel subunit alpha 1B-1.
 KW Calcium channel subunit; antagonist; agonist; diagnosis;
 KW Lambert Eaton Syndrome.
 OS Homo sapiens.
 PN WQ9504822-A.
 PD 16-FEB-1995.
 PF 11-AUG-1994; U09230.
 PR 11-AUG-1993; US-105536.
 PR 05-NOV-1993; US-149097.
 PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
 PI Ellis SB, Gillespie A, Harpold MM, McCue AF, Williams ME;
 DR WPI; 95-090900/12.
 DR N-PSDB; Q84657.
 PT DNA encoding human calcium channel sub-unit(s) - used for
 PT developing agonists and antagonists
 PT obtaining agonists and antagonists
 PS Disclosure; Page 138-149; 285pp; English.
 CC DNA encoding the alpha 1B subunit was isolated by screening a
 CC human basal ganglia cDNA library with fragments of the rabbit
 CC skeletal muscle calcium channel alpha 1 subunit-encoding cDNA.
 CC A portion of one of the positive clones was used to screen an
 CC IMR32 cell cDNA library. Clones that hybridized to the basal
 CC ganglia probe were used to further screen an IMR32 cell cDNA
 CC library to identify overlapping clones that in turn were used
 CC to screen a human hippocampus cDNA library. A series of clones
 CC to span nearly the entire length of the nt. sequence encoding
 CC the human alpha 1B subunit was obtained. Nucleic acid amplification
 CC of specific regions of the IMR32 cell alpha 1B mRNA yielded
 CC additional segments of the alpha 1B coding sequence. A full-
 CC length alpha 1B DNA clone was constructed by ligating portions
 CC of the partial cDNA clones. Nucleic acid amplification analysis
 CC of IMR32 cell RNA and genomic DNA using oligo primers correspond
 CC to sequences located 5' and 3' of the stop codon of the DNA encoding
 CC the alpha 1B subunit revealed an alternatively spliced alpha
 CC 1B-encoding mRNA in IMR32 cells. This second mRNA product is the
 CC result of differential splicing of the alpha 1B subunit transcript
 CC to include another exon that is not present in the mRNA correspond
 CC to the other 3' alpha 1B cDNA sequence that was initially isolated.
 CC The alpha 1B subunit encoded by a DNA sequence contig. an additional
 CC exon is referred to as alpha 1B-1 and given in Q84657/R71005,
 CC whereas the other form is referred to as alpha 1B-2 and is given in
 CC Q84658/R71006. Following the sequence of the additional exon in
 CC alpha 1B-1 the alpha 1B-1 and alpha 1B-2 sequences are identical.
 CC Sequence 2339 AA;

Query Match 85.4%; Score 35; DB 14; Length 2339;
 Best Local Similarity 83.3%; Pred. No. 2.06e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1985 pdgepq 1990
 | ||||
 Qy 3 PTGEPQ 8

RESULT 11
 ID W32784 standard; Protein; 59 AA.
 AC W32784;
 DT 03-FEB-1998 (first entry)

DE Rat thymus 60K-CSF protein.
 KW Rat; thymus; growth differentiation activity; thymic macrophage cell;
 OS cerebral microglial cell; cloning.
 PN Rattus rattus.
 J09248192-A.
 PD 22-SEP-1997.
 PF 26-JUL-1996; 214106.
 PR 09-JAN-1996; JP-018027.
 PA (MEIP) MEIJI MILK PROD CO LTD.
 DR WPI; 97-520745/48.
 DR N-PSDB; T92466.
 PT Thymus 60K-CSF protein DNA - encodes a protein having growth
 PT differentiation activity on thymic macrophage cells and cerebral
 PT microglial cells
 PS Claim 2; Page 4-5; 8pp; Japanese.
 CC The present sequence represents rat thymus 60K-CSF. The DNA has an
 CC approximate molecular weight of 60000 dalton by SDS-PAGE. The
 CC 60K-CSF protein induces the growth and differentiation of thymus
 CC macrophage cells and cerebral microglial cells. The DNA fragment
 CC can be used for cloning cDNA and genomic DNA corresponding to the
 CC 60K-CSF protein.
 SQ Sequence 59 AA;
 Query Match 82.9%; Score 34; DB 25; Length 59;
 Best Local Similarity 100.0%; Pred. No. 2.80e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 25 ptgsep 29
 QY 3 PTGEP 7
 RESULT 12
 ID W02134 standard; Protein; 109 AA.
 AC W02134;
 DT 09-APR-1997 (first entry)
 DE High mobility group HMGI-C.
 KW High mobility group; HMGI-C; probe; antisense; multi-tumour;
 KW aberrant growth gene; in situ; diagnosis; treatment;
 KW mesenchymal hamartoma; lipoma; atherosclerotic plaque;
 KW pleomorphic salivary gland endometrial polyp; tumour; benign;
 KW malignant; sarcoma; rhabdomyosarcoma; osteosarcoma; carcinoma;
 KW antibody; detection; hyperproliferation.
 OS Homo sapiens.
 PN EP-727487-A1.
 PD 21-AUG-1996.
 PF 14-JUL-1995; 201951.
 PR 17-FEB-1995; EP-200390.
 PA (LEUV-) LEUVEN RES & DEV.
 PI Bullerdielck J, Moils R, Schoenmakers HPPM, Van De Ven WJM;
 DR WPI; 96-372875/38.
 DR N-PSDB; T36365.
 PT Multi-tumour aberrant growth gene - and probes derived from it,
 PT useful in in situ diagnostic method for detecting tumour cells, e.g.
 PT mesenchymal hamartomas and rhabdomyosarcomas
 PS Claim 6; Pages 56-59; 78pp; English.
 CC The present sequence is protein product of the high mobility group
 CC (HMG) HMGI-C gene sequence (map position 12q15). Probes and
 CC antisense sequences derived from the HMGI-C gene, a multi-tumour
 CC aberrant growth (WAG) gene, can be useful in the in situ diagnosis
 CC and treatment of mesenchymal hamartomas (e.g. breast or lung),
 CC lipomas, pleomorphic salivary gland endometrial polyps,
 CC atherosclerotic plaques and other benign tumours, as well as
 CC malignant sarcomas (e.g. rhabdomyosarcomas and osteosarcomas) and
 CC carcinomas (e.g. of the breast, lung, skin or thyroid). The HMGI-C
 CC protein product, can be used for the production of antibodies for
 CC the detection of hyperproliferating cells, e.g. tumour cells.
 SQ Sequence 109 AA;
 Query Match 82.9%; Score 34; DB 20; Length 109;
 Best Local Similarity 100.0%; Pred. No. 2.80e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 25 ptgsep 29
 QY 3 PTGEP 7
 RESULT 13
 ID R81430 standard; Protein; 177 AA.
 AC R81430;
 DT 02-JUL-1996 (first entry)
 DE Hepatitis GB virus (HGBV) clone 16 protein prod.
 KW Hepatitis GB virus; HGBV; diagnosis; treatment; vaccine;
 KW reagents; non-A; non-B; non-C; non-D; non-E; clone 16;
 KW tamarin; infected plasma; lambda phage; cDNA library.
 OS Hepatitis GB virus.
 FH Key Location/Qualifiers
 FT misc_difference 58
 FT misc_difference /note= "corresponding codon STOP codon"
 FT misc_difference 63
 FT misc_difference /note= "corresponding codon STOP codon"
 FT misc_difference 165
 FT misc_difference /note= "corresponding codon STOP codon"
 PN W09521922-A2.
 PD 17-AUG-1995.
 PF 14-FEB-1995; U02118.
 PR 14-FEB-1994; US-196030.
 PR 13-MAY-1994; US-242654.
 PR 29-JUL-1994; US-283314.
 PR 23-NOV-1994; US-344190.
 PR 23-NOV-1994; US-344185.
 PR 27-JAN-1995; US-344557.
 PA (ABBO) ABBOTT LAB.
 PI Buijk SL, Dawson GJ, Desai SM, Erker JC, Leary TP;
 PI Muerhoff AS, Mushanwar IK, Pilot-Matias TJ, Schlander GG;
 PI Simons JN;
 DR WPI; 95-293123/38.
 DR N-PSDB; T00046.
 PT Non-A, non-B, non-C, non-D, non-E Hepatitis virus reagents - useful
 PT for diagnosis and therapy of hepatitis GB virus
 PS Example 5; Pages 208-209; 66pp; English.
 CC Double stranded hepatitis GB virus (HGBV) DNA obtd. from HGBV
 CC infected tamarin plasma, using standard procedures, was used to
 CC prepare a lambda phage HGBV cDNA library. The cDNA clone T00046,
 CC which encodes the proteins R81429-34 (the 6 possible reading
 CC frames), was rescued from the lambda phage, searched against a
 CC sequence database and found to be an unique HGBV sequence.
 CC Reagents which comprise the HGBV DNA, or its protein prods. can
 CC be used for the diagnosis, therapy or in a vaccine to prevent
 CC HGBV infection.
 SQ Sequence 177 AA;
 Query Match 82.9%; Score 34; DB 16; Length 177;
 Best Local Similarity 100.0%; Pred. No. 2.80e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 157 ptgsep 161
 QY 3 PTGEP 7
 RESULT 14
 ID W02132 standard; Protein; 343 AA.
 AC W02132;
 DT 09-APR-1997 (first entry)
 DE High mobility group protein:lipoma preferred partner fusion protein.
 KW High mobility group; lipoma preferred partner; fusion protein;
 KW DNA binding domain; HMGI-C; cysteine rich; zinc binding domain;
 KW LIM domain; translocation; tumour; mesenchymal; gene expression;
 KW aberrant; growth control.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT protein 1..60
 FT /note= "high mobility group protein fragment"
 FT protein 61..343

FT misc_difference 302 /note= "lipoma preferred partner fragment"
 FT /note= "corresponding codon TAG"
 FT misc_difference 320 /note= "corresponding codon TAA"
 FT misc_difference 325 /note= "corresponding codon TAA"
 FT EP-727487-A1.
 PN 21-AUG-1996.
 PD 14-JUL-1995; 201951.
 PR 17-FEB-1995; EP-200390.
 PA (LEUV-) LEUVEN RES & DEV.
 PI Bullerdiek J, MoIs R, Schoenmakers HPPM, Van De Ven WJN;
 DR WPI; 96-372875/38.
 DR N-PSDB: T36363.
 PT Multi-tumour aberrant growth gene - and probes derived from it,
 PT useful in in situ diagnostic method for detecting tumour cells, e.g.
 PT mesenchymal hamartomas and rhabdomyosarcomas
 FS Example 3; Pages 45-47; 78pp; English.
 CC The present sequence is the predicted protein product of a 3'-RACE
 CC product, comprising the junction region of a high mobility group
 CC protein (HMG):lipoma preferred partner (LPP) fusion gene. The
 CC protein product was predicted to contain 3 DNA binding domains
 CC (DBD) of the HMG1-C (map position 12q15), and the 3 cysteine rich
 CC zinc binding (LIM) domains of the LPP (map position 3q27-28). In
 CC PCR analysis using appropriate nested amplicon sets, similar
 CC HMG:LPP hybrid transcripts were detected in various primary lipomas
 CC and lipoma cell lines carrying a chromosome 3 to 12 translocation
 CC (t(3;12)), and in a cytogenetically normal lipoma. This revealed
 CC that the cytogenetically detectable, and hidden, t(3;12) in lipomas
 CC result in the in-phase fusion of the DBD of HMG1-C to the LIM
 CC domains of LPP. Consequently, these LIM domains are most likely
 CC presented in the nuclear environment of these lipoma cells, where
 CC they might affect gene expression, possibly leading to aberrant
 CC growth control. This in an example of a benign mesenchymal tumour
 CC with chromosome 12q13-q15 aberrations, where a chromosome
 CC translocation partner contributes recurrently and consistently to
 CC the formation of a well-defined tumour associated HMG1-C fusion
 CC protein.
 SQ Sequence 343 AA;

Query Match 82.9%; Score 34; DB 20; Length 343;
 Best Local Similarity 100.0%; Pred. No. 2.80e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 16 ptgep 20
 |||||
 Qy 3 PTGEP 7

RESULT 15
 ID W38218 standard; Protein; 649 AA.
 AC W38218;
 DT 22-MAY-1998 (first entry)
 DE Maize starch synthase type I.
 KW Maize; starch synthase type I; starch.
 OS Zea mays.
 PN DE19619918-AL.
 PD 20-NOV-1997.
 PF 17-MAY-1996; 019918.
 PR 17-MAY-1996; DE-019918.
 PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.
 PI Froberg C, Kossmann J;
 DR WPI; 98-000821/01.
 DR N-PSDB: T95785.
 PT DNA encoding maize starch synthase type I protein - for producing
 PT transgenic plants
 PS Claim 1; Pages 16-20; 23pp; German.
 CC The present sequence is maize starch synthase type I, useful
 CC in the production of starch. Starch can be used in various
 CC conventional starch applications, e.g. starch hydrolysate products,
 CC foods, papermaking, adhesives, textiles, building materials, soil
 CC stabilisation, agrochemicals, pharmaceuticals, cosmetics, coal

CC briquettes, ore and coal slurries, foundry casting, rubber, leather
 CC and synthetic polymers. The enzyme produces a starch stated to have
 CC different physicochemical properties, especially viscosity and
 CC gelling properties, from wild type starch.
 SQ Sequence 649 AA;

Query Match 82.9%; Score 34; DB 29; Length 649;
 Best Local Similarity 100.0%; Pred. No. 2.80e+02;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 83 ptgep 87
 |||||
 Qy 3 PTGEP 7

Search completed: Thu May 13 15:40:16 1999
 Job time : 17 secs.

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run On: Thu May 13 15:41:05 1999; MasPar time 1.94 Seconds

Tabular output not generated. 83.113 Million cell updates/sec

Title: >US-09-040-485-8

Description: (3-8) from US09040485.pep (2 of 2)

Perfect Score: 41

Sequence: 1 PTGEPQ 6

Scoring table:

PAM 150

Gap 15

Searched: 74019 seqs, 26840295 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot35

1:swissprot

Statistics: Mean 18.021; Variance 16.071; scale 1.121

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	41	100.0	757	1	ASPH_HUMAN	5.09e-01
2	38	92.7	232	1	ASPARTYL/ASPARAGINYL B	3.60e+00
3	38	92.7	485	1	HYPOTHETICAL GENE 8 ME	3.60e+00
4	38	92.7	518	1	GLUTAMYL-TRNA SYNTHETA	3.60e+00
5	38	92.7	520	1	PUTATIVE CYTOCHROME P4	3.60e+00
6	38	92.7	520	1	PUTATIVE CYTOCHROME P4	3.60e+00
7	37	90.2	116	1	PUTATIVE CYTOCHROME P4	3.60e+00
8	37	90.2	301	1	PROTEIN-TYROSINE PHOSP	6.75e+00
9	37	90.2	325	1	VP6 PROTEIN (MINOR INN	6.75e+00
10	37	90.2	325	1	VP6 PROTEIN (MINOR INN	6.75e+00
11	37	90.2	325	1	VP6 PROTEIN (MINOR INN	6.75e+00
12	37	90.2	325	1	VP6 PROTEIN (MINOR INN	6.75e+00
13	37	90.2	328	1	VP6 PROTEIN (MINOR INN	6.75e+00
14	37	90.2	329	1	VP6 PROTEIN (MINOR INN	6.75e+00
15	37	90.2	370	1	VP6 PROTEIN (MINOR INN	6.75e+00
16	37	90.2	468	1	GLUTAMYL-TRNA SYNTHETA	6.75e+00
17	36	87.8	142	1	GENERAL SECRETION PATH	1.25e+01
18	36	87.8	612	1	GSPG_PSEAE	1.25e+01
19	36	87.8	1079	1	GTP-BINDING PROTEIN LE	1.25e+01
20	35	85.4	218	1	YD23 SCHPO	1.25e+01
21	35	85.4	231	1	POU2 XENLA	2.29e+01
22	35	85.4	251	1	RNC_SYNY3	2.29e+01
23	35	85.4	311	1	PRP2_HUMAN	2.29e+01
					SEX-DETERMINING REGION	2.29e+01

24	35	85.4	331	1	PRP1_HUMAN	SALIVARY PROLINE-RICH	2.29e-01
25	35	85.4	333	1	CAP1_DICDI	CAMP BINDING PROTEIN C	2.29e-01
26	35	85.4	337	1	GALE_SALTY	UDP-GLUCOSE 4-EPIMERAS	2.29e-01
27	35	85.4	355	1	SRY_MUSSP	SEX-DETERMINING REGION	2.29e-01
28	35	85.4	512	1	FUS1_YEAST	NUCLEAR FUSION PROTEIN	2.29e-01
29	35	85.4	535	1	YZ25_MYCTU	HYPOTHETICAL 56.2 KD P	2.29e-01
30	35	85.4	545	1	HMW_DESVH	HIGH-MOLECULAR-WEIGHT	2.29e-01
31	35	85.4	562	1	O16G_BACTR	OLIGO-1,6-GLUCOSIDASE	2.29e-01
32	35	85.4	618	1	DNK_DEIPR	DNK PROTEIN (HEAT SHO	2.29e-01
33	35	85.4	622	1	MAK_RAT	SERINE/THREONINE-PROTE	2.29e-01
34	35	85.4	622	1	MAK_MOUSE	SERINE/THREONINE-PROTE	2.29e-01
35	35	85.4	2336	1	CIC5_RAT	N-TYPE CALCIUM CHANNEL	2.29e-01
36	35	85.4	2339	1	CIC5_HUMAN	N-TYPE CALCIUM CHANNEL	2.29e-01
37	35	85.4	3712	1	LMA_DROME	LAMININ ALPHA CHAIN PR	2.29e-01
38	34	82.9	296	1	PRP3_MOUSE	PROLINE-RICH PROTEIN M	4.12e-01
39	34	82.9	369	1	PSPD_BOVIN	PULMONARY SURFACTANT-A	4.12e-01
40	34	82.9	381	1	YE2I_CAEEL	HYPOTHETICAL 42.2 KD P	4.12e-01
41	34	82.9	475	1	TRB1_ECOLI	TRAB PROTEIN.	4.12e-01
42	34	82.9	491	1	PRCC_HUMAN	PROLINE-RICH PROTEIN P	4.12e-01
43	34	82.9	928	1	YL83_CAEEL	HYPOTHETICAL 105.8 KD	4.12e-01
44	34	82.9	1220	1	ATCP_PIG	CALCIUM-TRANSPORTING A	4.12e-01
45	34	82.9	1319	1	MNI_HUMAN	PROBABLE TUMOR SUPPRES	4.12e-01

ALIGNMENTS

RESULT 1
ID ASPH_HUMAN STANDARD; PRT; 757 AA.
AC Q12797;

DT 01-NOV-1997 (REL. 35, CREATED)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)

DE ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-DE HYDROXYLASE). (ASP BETA-HYDROXYLASE) (PEPTIDE-ASPARTATE BETA-DE DIOXYGENASE).

GN ASPH

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 95121937.

RA KORIOTH F., GIEFFERS C., FREY J.;

RL GENE 150:395-399(1994).

CC -!- FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF PROTEINS.

CC -!- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) -> PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).

CC -!- COFACTOR: IRON.

CC -!- SUBUNIT: MONOMER (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM.

CC -!- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES TESTED.

CC -!- PTM: MIGHT BE PROCESSED TO THE 56 KD (AA 274-757) OR 52 KD (AA 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY SIMILARITY).

EMBL: U03109; E82591; -

DR MIM: 600582; -

KW OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR; ENDOPLASMIC RETICULUM.

FT DOMAIN 1 54

FT TRANSMEM 55 75

FT DOMAIN 76 757

FT DOMAIN 13 20

FT DOMAIN 323 332

FT CARBOHYD 452 452

FT CARBOHYD 705 705

FT SEQUENCE 757 AA; 85498 MW; AE6AFC24 CRC32;

FT CYTOPLASMIC (POTENTIAL).

FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).

FT LUMENAL (POTENTIAL).

FT POLY-SER.

FT POLY-LYS.

FT POTENTIAL.

FT POTENTIAL.

Query Match 100.0%; Score 41; DB 1; Length 757;
Best Local Similarity 100.0%; Pred. No. 5.09e-01;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 175 PTGEPQ 180
|||||
QY 3 PTGEPQ 8

RESULT 2
ID VG08_HSV11 STANDARD; PRT; 232 AA.
AC Q00137;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL GENE 8 MEMBRANE PROTEIN.
GN 8.
OS ICTALURID HERPESVIRUS 1 (CHANNEL CATFISH VIRUS) (CCV).
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AUBURN 1;
RX MEDLINE; 92087490.
RA DAVISON A.J.;
RL VIROLOGY 186:9-14(1992).
DR EMBL; M75136; G331295; -.
DR EMBL; M75136; G331218; -.
DR PIR; I36786; MMBE13.
KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
FT TRANSMEM 166 182 POTENTIAL.
FT TRANSMEM 195 211 POTENTIAL.
SQ SEQUENCE 232 AA; 25466 MW; AA90E31E CRC32;

Query Match 92.7%; Score 38; DB 1; Length 232;
Best Local Similarity 83.3%; Pred. No. 3.60e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 90 PTGEPQ 95
|||||
QY 3 PTGEPQ 8

RESULT 3
ID SYE_RHIME STANDARD; PRT; 485 AA.
AC P15189;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE GLUTAMYL-TRNA SYNTHETASE (EC 6.1.1.17) (GLUTAMATE--TRNA LIGASE)
DE (GLURS).
GN GLTX.
OS RHIZOBIUM MELILOTI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC RHIZOBIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A2;
RX MEDLINE; 89291743.
RA LABERGE S.; GAGNON Y.; BORDELEAU L.M.; LAPOINTE J.;
RL J. BACTERIOL. 171:3926-3932(1989).
CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + TRNA(GLU) = AMP +
CC PYROPHOSPHATE + L-GLUTAMYL-TRNA(GLU).
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
DR EMBL; M27221; G717082; -.
DR PIR; A32888; SYRZET.
DR PROSITE; PS00178; AA-TRNA-LIGASE-I; 1.
KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
FT SIMILAR 12 22 "HIGH" REGION.
FT SIMILAR 253 257 "KMSK" REGION.
FT BINDING 256 256 ATP (BY SIMILARITY).
SQ SEQUENCE 485 AA; 54356 MW; 22131735 CRC32;

Query Match 92.7%; Score 38; DB 1; Length 485;

Best Local Similarity 83.3%; Pred. No. 3.60e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 14 PTGEPH 19
|||||
QY 3 PTGEPQ 8

RESULT 4
ID YRVA_CAEEL STANDARD; PRT; 518 AA.
AC Q27519;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PUTATIVE CYTOCHROME P450 T10B9.10 IN CHROMOSOME II (EC 1.14.-.-).
GN T10B9.10.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA GARDNER A.;
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; Z48717; G732602; -.
DR WORMPEP; T10B9.10; CE01655.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW HYPOTHETICAL PROTEIN; OXIDOREDUCTASE; MONOOXYGENASE; HEME.
FT BINDING 464 464 HEME (BY SIMILARITY).
SQ SEQUENCE 518 AA; 58999 MW; F66B1048 CRC32;

Query Match 92.7%; Score 38; DB 1; Length 518;
Best Local Similarity 83.3%; Pred. No. 3.60e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 270 PSGEQ 275
|||||
QY 3 PTGEPQ 8

RESULT 5
ID YRV2_CAEEL STANDARD; PRT; 520 AA.
AC Q27514;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PUTATIVE CYTOCHROME P450 T10B9.2 IN CHROMOSOME II (EC 1.14.-.-).
GN T10B9.2.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA GARDNER A.;
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; Z48717; G732596; -.
DR WORMPEP; T10B9.2; CE01656.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW HYPOTHETICAL PROTEIN; OXIDOREDUCTASE; MONOOXYGENASE; HEME.
FT BINDING 464 464 HEME (BY SIMILARITY).
SQ SEQUENCE 520 AA; 59524 MW; 087B5D9B CRC32;

Query Match 92.7%; Score 38; DB 1; Length 520;
Best Local Similarity 83.3%; Pred. No. 3.60e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Db 270 PSCEPQ 275
  1:||||
QY 3 PTGEPQ 8

RESULT 6
ID XRV1_CAEEL STANDARD; PRT; 520 AA.
AC Q27513;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PUTATIVE CYTOCHROME P450 T10B9.1 IN CHROMOSOME II (EC 1.14.-.-).
GN T10B9.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA GARDNER A.;
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; Z48717; G732595; -.
DR WORMPEP; T10B9.1; CE01654.
DR PROSITE; PS00086; CYTOCHROME P450; 1.
KW HYPOTHETICAL PROTEIN: OXIDOREDUCTASE; MONOOXYGENASE; HEME.
FT BINDING 464 464
SQ SEQUENCE 520 AA; 59371 MW; CFAVD0F8 CRC32;

Query Match 92.7%; Score 38; DB 1; Length 520;
Best Local Similarity 83.3%; Pred. No. 3.60e+00;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 270 PSCEPQ 275
  1:||||
QY 3 PTGEPQ 8

RESULT 7
ID PT10_STYPL STANDARD; PRT; 116 AA.
AC P28202;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE PROTEIN-TYROSINE PHOSPHATASE 10 (EC 3.1.3.48) (FRAGMENT).
GN STY 10.
OS STYELA PLICATA (SEA SQUIRT).
OC EUKARYOTA; METAZOA; CHORDATA; TUNICATA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91139172.
RA MATTHEWS R.J., FLORES E., THOMAS M.L.;
RL IMMUNOGENETICS 33:33-41(1991).
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
DR EMBL; M37995; G160882; -.
DR HSSP; P18031; 2HNP.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; PARTIAL.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; PARTIAL.
DR PROSITE; PS50055; TYR_PHOSPHATASE_FTP; 1.
KW HYDROLASE.
FT NON_TER 1
FT NON_TER 116
SQ SEQUENCE 116 AA; 13251 MW; F7568322 CRC32;

Query Match 90.2%; Score 37; DB 1; Length 116;
Best Local Similarity 66.7%; Pred. No. 6.75e+00;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 31 PSQDPQ 36
  1:||||

```

```

QY 3 PTGEPQ 8

RESULT 8
ID VP6_BTVA STANDARD; PRT; 301 AA.
AC P33422;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE VP6 PROTEIN (MINOR INNER CORE PROTEIN VP6).
GN S9.
OS BLUETONGUE VIRUS (SEROTYPE 2 / ISOLATE USA).
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93033709.
RA HWANG G.-Y., CHIOU J.-F., YANG Y.-Y., LI J.K.-K.;
RL VIRUS RES. 24:315-323(1992).
CC -!- FUNCTION: SURROUNDS AND INTERACTS WITH THE GENOMIC DS-RNA.
CC POSSESSES SS- AND DS-RNA-BINDING CAPACITY. ITS HYDROPHILIC
CC NATURE AND CAPABILITY TO BIND SS- AND DS-RNA SUGGEST ITS
CC INTERACTIONS WITH BTV GENOMIC RNA.
CC -!- SUBCELLULAR LOCATION: INNER CAPSID.
DR EMBL; L08668; G210851; -.
KW CORE PROTEIN.
FT DOMAIN 60 102 GLY-RICH.
SQ SEQUENCE 301 AA; 32864 MW; B21871B4 CRC32;

Query Match 90.2%; Score 37; DB 1; Length 301;
Best Local Similarity 66.7%; Pred. No. 6.75e+00;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 261 PTGDPH 266
  1:||||
QY 3 PTGEPQ 8

RESULT 9
ID VP6_BTVA STANDARD; PRT; 325 AA.
AC P32933;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE VP6 PROTEIN (MINOR INNER CORE PROTEIN VP6).
GN S9.
OS BLUETONGUE VIRUS (SEROTYPE 11 / ISOLATE USA).
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93033709.
RA HWANG G.-Y., CHIOU J.-F., YANG Y.-Y., LI J.K.-K.;
RL VIRUS RES. 24:315-323(1992).
CC -!- FUNCTION: SURROUNDS AND INTERACTS WITH THE GENOMIC DS-RNA.
CC POSSESSES SS- AND DS-RNA-BINDING CAPACITY. ITS HYDROPHILIC
CC NATURE AND CAPABILITY TO BIND SS- AND DS-RNA SUGGEST ITS
CC INTERACTIONS WITH BTV GENOMIC RNA.
CC -!- SUBCELLULAR LOCATION: INNER CAPSID.
DR EMBL; L08670; G210845; -.
DR PIR; A48561; A48561.
KW CORE PROTEIN.
FT DOMAIN 84 125 GLY-RICH.
SQ SEQUENCE 325 AA; 35371 MW; 869B26B3 CRC32;

Query Match 90.2%; Score 37; DB 1; Length 325;
Best Local Similarity 66.7%; Pred. No. 6.75e+00;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 285 PTGDPH 290
  1:||||
QY 3 PTGEPQ 8

RESULT 10

```

ID VP6_BT17 STANDARD; PRT; 325 AA.
 AC P32935;
 DT 01-OCT-1993 (REL. 27, CREATED)
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
 DE VP6 PROTEIN (MINOR INNER CORE PROTEIN VP6).
 GN S9.
 OS BLUETONGUE VIRUS (SEROTYPE 17 / ISOLATE USA).
 OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93033709.
 RA HWANG G.-Y., CHIOU J.-F., YANG Y.-Y., LI J.K.-K.;
 RL VIRUS RES. 24:315-323(1992).
 CC -!- FUNCTION: SURROUNDS AND INTERACTS WITH THE GENOMIC DS-RNA.
 CC POSSESSES SS- AND DS-RNA-BINDING CAPACITY. ITS HYDROPHILIC
 CC NATURE AND CAPABILITY TO BIND SS- AND DS-RNA SUGGEST ITS
 CC INTERACTIONS WITH BTV GENOMIC RNA.
 CC -!- SUBCELLULAR LOCATION: INNER CAPSID.
 DR EMBL; L08672; G210849; -.
 KW CORE PROTEIN.
 FT DOMAIN 84 125 GLY-RICH.
 SQ SEQUENCE 325 AA; 5D26438B CRC32;

Query Match 90.2%; Score 37; DB 1; Length 325;
 Best Local Similarity 66.7%; Pred. No. 6.75e+00;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 285 PTGDPH 290
 |||:|
 QY 3 PTGEPQ 8

RESULT 11
 ID VP6_BT10 STANDARD; PRT; 325 AA.
 AC P33423;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
 DE VP6 PROTEIN (MINOR INNER CORE PROTEIN VP6) (VERSION 2).
 GN S9.
 OS BLUETONGUE VIRUS (SEROTYPE 10 / ISOLATE USA).
 OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93033709.
 RA HWANG G.-Y., CHIOU J.-F., YANG Y.-Y., LI J.K.-K.;
 RL VIRUS RES. 24:315-323(1992).
 CC -!- FUNCTION: SURROUNDS AND INTERACTS WITH THE GENOMIC DS-RNA.
 CC POSSESSES SS- AND DS-RNA-BINDING CAPACITY. ITS HYDROPHILIC
 CC NATURE AND CAPABILITY TO BIND SS- AND DS-RNA SUGGEST ITS
 CC INTERACTIONS WITH BTV GENOMIC RNA.
 CC -!- SUBCELLULAR LOCATION: INNER CAPSID.
 DR EMBL; L08669; G210843; -.
 KW PIR; B48561; B48561.
 RN CORE PROTEIN.
 FT DOMAIN 85 125 GLY-RICH.
 SQ SEQUENCE 325 AA; 1141118E CRC32;

Query Match 90.2%; Score 37; DB 1; Length 325;
 Best Local Similarity 66.7%; Pred. No. 6.75e+00;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 285 PTGDPH 290
 |||:|
 QY 3 PTGEPQ 8

RESULT 12
 ID VP6_BT13 STANDARD; PRT; 325 AA.
 AC P32934;
 DT 01-OCT-1993 (REL. 27, CREATED)

DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
 DE VP6 PROTEIN (MINOR INNER CORE PROTEIN VP6).
 GN S9.
 OS BLUETONGUE VIRUS (SEROTYPE 13 / ISOLATE USA).
 OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93033709.
 RA HWANG G.-Y., CHIOU J.-F., YANG Y.-Y., LI J.K.-K.;
 RL VIRUS RES. 24:315-323(1992).
 CC -!- FUNCTION: SURROUNDS AND INTERACTS WITH THE GENOMIC DS-RNA.
 CC POSSESSES SS- AND DS-RNA-BINDING CAPACITY. ITS HYDROPHILIC
 CC NATURE AND CAPABILITY TO BIND SS- AND DS-RNA SUGGEST ITS
 CC INTERACTIONS WITH BTV GENOMIC RNA.
 CC -!- SUBCELLULAR LOCATION: INNER CAPSID.
 DR EMBL; L08671; G210847; -.
 KW CORE PROTEIN.
 FT DOMAIN 84 125 GLY-RICH.
 SQ SEQUENCE 325 AA; 7F5B6F92 CRC32;

Query Match 90.2%; Score 37; DB 1; Length 325;
 Best Local Similarity 66.7%; Pred. No. 6.75e+00;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 285 PTGDPH 290
 |||:|
 QY 3 PTGEPQ 8

RESULT 13
 ID VP6_BT15 STANDARD; PRT; 328 AA.
 AC Q03328; Q65752;
 DT 01-OCT-1993 (REL. 27, CREATED)
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE VP6 PROTEIN (MINOR INNER CORE PROTEIN VP6).
 GN S9.
 OS BLUETONGUE VIRUS (SEROTYPE 1 / ISOLATE SOUTH AFRICA).
 OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93057380.
 RA WADE-EVANS A.M., MERTENS P.P.C., BELSHAM G.J.;
 RL J. GEN. VIROL. 73:3023-3026(1992).
 CC -!- FUNCTION: SURROUNDS AND INTERACTS WITH THE GENOMIC DS-RNA.
 CC POSSESSES SS- AND DS-RNA-BINDING CAPACITY. ITS HYDROPHILIC
 CC NATURE AND CAPABILITY TO BIND SS- AND DS-RNA SUGGEST ITS
 CC INTERACTIONS WITH BTV GENOMIC RNA.
 CC -!- SUBCELLULAR LOCATION: INNER CAPSID.
 DR EMBL; D10905; G221081; -.
 KW PIR; JQ1875; JQ1875.
 RN CORE PROTEIN.
 FT CHAIN 1 328 VP6.
 FT CHAIN 5 328 VP6A.
 FT DOMAIN 88 130 GLY-RICH.
 SQ SEQUENCE 328 AA; 35928 MW; D48A20F2 CRC32;

Query Match 90.2%; Score 37; DB 1; Length 328;
 Best Local Similarity 66.7%; Pred. No. 6.75e+00;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294
 |||:|
 QY 3 PTGEPQ 8

RESULT 14
 ID VP61_BT10 STANDARD; PRT; 329 AA.
 AC P23066;
 DT 01-NOV-1991 (REL. 20, CREATED)
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)

DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE VP6 PROTEIN (MINOR INNER CORE PROTEIN VP6) (VERSION 1).
GN S9.
OS BLUFONGUE VIRUS (SEROTYPE 10 / ISOLATE USA).
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89293076.
RA FUKUSHO A., YU Y., YAMAGUCHI S., ROY P.;
RL J. GEN. VIROL. 70:1677-1689(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90345726.
RA ROY P., MARSHALL J.J.A., FRENCH T.J.;
RL CURR. TOP. MICROBIOL. IMMUNOL. 162:43-87(1990).
RN [3]
RP REVISIONS.

RA BLACHERE C.;
RL SUBMITTED (JAN-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: SURROUNDS AND INTERACTS WITH THE GENOMIC DS-RNA.
CC POSSESSES SS- AND DS-RNA-BINDING CAPACITY. ITS HYDROPHILIC
CC NATURE AND CAPABILITY TO BIND SS- AND DS-RNA SUGGEST ITS
CC INTERACTIONS WITH BTV GENOMIC RNA.
CC -!- SUBCELLULAR LOCATION: INNER CAPSID.
CC -!- CAUTION: SEE ALSO VERSION 2 OF THIS PROTEIN.
DR EMBL; D00509; G221077; -.
DR PIR; B32400; VEXRG3.
DR PIR; S10542; S10542.
KW CORE PROTEIN.88 130 GLY-RICH.
FT DOMAIN
SQ SEQUENCE 329 AA; 35515 MW; 701CD715 CRC32;

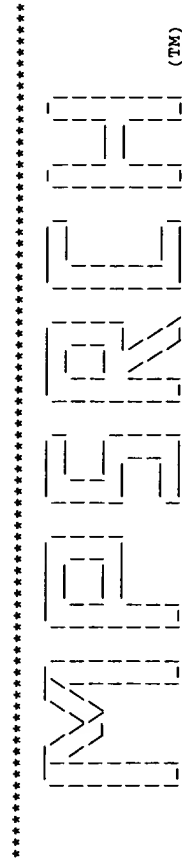
Query Match 90.2%; Score 37; DB 1; Length 329;
Best Local Similarity 66.7%; Pred. No. 6.75e+00;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294
|||:|
QY 3 PTGEPQ 8

RESULT 15
ID SVE_CHLPS STANDARD; PRT; 370 AA.
AC Q06560;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE GLUTAMYL-TRNA SYNTHETASE (EC 6.1.1.17) (GLUTAMATE--TRNA LIGASE)
DE (GLURS) (FRAGMENT).
GN GLTX.
OS CHLAMYDIA PSITTACI.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; RICKETTSIAS; CHLAMYDIALES;
OC CHLAMYDIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6BC;
RX MEDLINE; 93259937.
RA WICHLAN D.W., HATCH T.P.;
RL J. BACTERIOL. 175:2936-2942(1993).
CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + TRNA(GLU) = AMP +
CC PYROPHOSPHATE + L-GLUTAMYL-TRNA(GLU).
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
DR EMBL; L13598; G289833; -.
DR PIR; A36909;
DR PROSITE; PS00178; AA-TRNA-LIGASE_I; 1.
KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
FT SIMILAR 14 24 "HIGH" REGION.
FT SIMILAR 255 259 "RMSKS" REGION.
FT BINDING 258 258 ATP (BY SIMILARITY).
FT NON-TER 370 370
SQ SEQUENCE 370 AA; 43093 MW; 5F53F41B CRC32;

Query Match 90.2%; Score 37; DB 1; Length 370;
Best Local Similarity 66.7%; Pred. No. 6.75e+00;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 16 PTGDPH 21
|||:|
QY 3 PTGEPQ 8

Search completed: Thu May 13 15:41:13 1999
Job time : 8 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 13 15:44:14 1999; MasPar time 2.54 Seconds
Tabular output not generated. 63.620 Million cell updates/sec

Title: >US-09-040-485-9
Description: (1-10) from US09040485.pep
Perfect Score: 61
Sequence: 1 QENPDSSEPV 10

Scoring table: PAM 150
Gap 15

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29

Statistics: Mean 14.163; Variance 44.856; scale 0.316

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description	Pred. No.
1	43	70.5	995	12 R60812	Agarase 0107 from Vib	1.67e+02
2	42	68.9	332	20 W06491	Beta-1-4-galactosyltr	2.13e+02
3	41	67.2	14	29 W40864	Cytotoxic Epstein-Bar	2.71e+02
4	41	67.2	84	13 R66800	N-terminal fragment (2.71e+02
5	41	67.2	154	13 R66801	N-terminal fragment (2.71e+02
6	41	67.2	188	13 R66802	N-terminal fragment (2.71e+02
7	41	67.2	228	13 R66804	N-terminal fragment (2.71e+02
8	41	67.2	289	13 R66796	Novel mouse proteogly	2.71e+02
9	41	67.2	311	29 W41156	Mouse syndecan protei	2.71e+02
10	41	67.2	311	13 R66793	Novel mouse proteogly	2.71e+02
11	41	67.2	311	10 R55276	Syndecan protein.	2.71e+02
12	41	67.2	311	17 R87001	Mouse syndecan-1.	2.71e+02
13	41	67.2	919	23 W18580	Potato alpha-glucosid	2.71e+02
14	40	65.6	10	27 W42736	Antigenic decapeptide	3.43e+02
15	40	65.6	10	28 W42824	Antigenic decapeptide	3.43e+02
16	40	65.6	151	27 W42642	Human sperm zona bind	3.43e+02
17	40	65.6	151	14 R79761	Human testis sperm zo	3.43e+02
18	40	65.6	162	28 W42692	Recombinant human Spi	3.43e+02

19	40	65.6	163	28 W42693	Baboon sperm zona bin	3.43e+02
20	40	65.6	331	26 W29817	Mammalian AMPK-gamma	3.43e+02
21	40	65.6	461	27 W33890	Flea saliva protein P	3.43e+02
22	39	63.9	226	3 R13901	Nitrile hydratase L t	4.34e+02
23	39	63.9	260	21 W14422	Human Dnase I variant	4.34e+02
24	39	63.9	496	22 W08605	Chimeric MOMLV and Ty	4.34e+02
25	39	63.9	517	22 W08607	Chimeric MOMLV and Ty	4.34e+02
26	39	63.9	866	19 W04185	Human interleukin-17	4.34e+02
27	39	63.9	2019	13 R67913	Cardiac sodium channe	4.34e+02
28	39	63.9	2020	2 R67913	Cardiac sodium channe	4.34e+02
29	38	62.3	147	27 W36001	Human Fhit protein.	5.48e+02
30	38	62.3	147	28 W39124	Human PSAFP-1 protein	5.48e+02
31	38	62.3	537	29 W46736	Amino acid sequence O	5.48e+02
32	38	62.3	591	29 W41418	Paxillin beta isoform	5.48e+02
33	38	62.3	605	29 W41419	Paxillin gamma isoform	5.48e+02
34	38	62.3	1100	1 P95644	Rabbit skeletal muscle	5.48e+02
35	38	62.3	1103	14 R71012	Human neuronal calciu	5.48e+02
36	38	62.3	1106	22 W18389	Rabbit calcium channe	5.48e+02
37	38	62.3	1106	13 R73056	Rabbit skeletal calci	5.48e+02
38	38	62.3	1106	26 W37712	Rabbit skeletal calci	5.48e+02
39	38	62.3	1203	8 R41668	Human endothelial cel	5.48e+02
40	38	62.3	1443	29 W44137	Homo sapiens factor V	5.48e+02
41	38	62.3	2133	29 W44133	Sus scrofa factor VII	5.48e+02
42	38	62.3	2986	8 R44432	eryA region polypepti	5.48e+02
43	37	60.7	234	3 P50139	Rabbit tumor necrosis	6.92e+02
44	37	60.7	552	3 P60956	Sequence of bovine me	6.92e+02
45	37	60.7	761	9 R48036	Mycobacterium BCG imm	6.92e+02

ALIGNMENTS

RESULT 1
ID R60812 standard; Protein; 995 AA.
AC R60812;
DT 20-JUL-1995 (first entry)
DE Agarase 0107 from Vibrio species.
KW Agarase; Vibrio; PCR; amplification; amplify; primer; probe; E.coli;
KW cation-exchange; chromatograph; neogaroooligosaccharide.
OS Vibrio sp. J70107.
FS Key Location/Qualifiers
FT peptide 1..20
FT protein /label= signal peptide
FT protein /label= mature peptide
FT J0628488-A.
FN 11-OCT-1994.
PD 01-APR-1993; 096549.
PR 01-APR-1993; JP-096549
PA (NIBS) JAPAN TOBACCO INC.
DR WPI; 94-362595/45.
DR N-PSDB; Q73755.
PT DNA encoding protein with agarase activity - useful for the
PT industrial production of a large amount of
PT neo-agarao-oligosaccharide
PS Claim 1; Page 8-12; 13pp; Japanese.
CC The sequence of a novel protein containing agarase activity from
CC Vibrio species. The gene of 3205 bp encodes a protein of 995 amino
CC acids. The gene was obtained by PCR amplification from a Vibrio genomic
CC library using primers and probes based on the amino acid sequence
CC (R60813-21) from protein sequencing of partially degraded agarase 0107.
CC The gene was then introduced into E.coli JM83. The expressed protein
CC could then be recovered by cation-exchange chromatography. The protein
CC can be used in the large scale manufacture of neogaroooligosaccharide.
SQ Sequence 995 AA;

Query Match 70.5%; Score 43; DB 12; Length 995;
Best Local Similarity 66.7%; Pred. No. 1.67e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 947 eapdpgepv 955
Qy 2 ENPDSSEPV 10

```

CC epitope.
SQ Sequence 14 AA;

Query Match 67.2%; Score 41; DB 29; Length 14;
Best Local Similarity 55.6%; Pred. No. 2.71e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 1 depasteqv 9
QY 2 ENPDSSEPV 10

RESULT 4
ID R66800 standard; Peptide; 84 AA.
AC R66800;
DT 11-SEP-1995 (first entry)
DE N-terminal fragment (a.a. 1-84) of mature mouse syndecan-1.
KW Cell surface; proteoglycan; syndecan; mouse; mammary; epithelial cell;
KW ectodomain; NMG; glycosylation; heparan sulphate; chondroitin sulphate;
KW glycosaminoglycan; chimaera; chimaeric molecule; effector molecule;
KW receptor; drug; antibody; diagnostic agent.
OS Mus musculus.
PN W09500633-A.
PD 05-JAN-1995.
PF 17-JUN-1994; U06920.
PR 17-JUN-1993; US-078683.
PA (CHIL-) CHILDRENS MEDICAL CENT.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Bernfield M, Kato M, Saunders S;
DR WPI; 95-052071/07.
PT DNA and protein sequences for recombinant syndecan-derived
PT proteoglycans - comprising a core protein having glycosylation
PT sites for heparin sulphate glycosaminoglycan side chains.
PS Claim 15; Page 82; 97pp; English.
CC Peptides R66797-802 correspond to varying length fragments of the
CC N-terminal sequence of the mature mouse syndecan-1 (see R66793 and
CC R66796). The sequence given here covers the amino acids from 1-84 of the
CC mature protein sequence. The fragments are portions of the extracellular
CC domain of the protein and contain a heparan sulphate glycosaminoglycan
CC attachment site (R66795). The functional domains, esp. the soluble
CC extracellular or heparan binding site, of the syndecan molecules (see
CC R66797-812 and R66818) can be used to construct chimaeras by linking them
CC to biological effector molecules, cell surface receptors, drugs,
CC antibodies, diagnostic agents or components of microorganisms.
SQ Sequence 84 AA;

Query Match 67.2%; Score 41; DB 13; Length 84;
Best Local Similarity 55.6%; Pred. No. 2.71e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 74 ekpegeqv 82
QY 2 ENPDSSEPV 10

RESULT 5
ID R66801 standard; Peptide; 154 AA.
AC R66801;
DT 11-SEP-1995 (first entry)
DE N-terminal fragment (a.a. 1-154) of mature mouse syndecan-1.
KW Cell surface; proteoglycan; syndecan; mouse; mammary; epithelial cell;
KW ectodomain; NMG; glycosylation; heparan sulphate; chondroitin sulphate;
KW glycosaminoglycan; chimaera; chimaeric molecule; effector molecule;
KW receptor; drug; antibody; diagnostic agent.
OS Mus musculus.
PN W09500633-A.
PD 05-JAN-1995.
PF 17-JUN-1994; U06920.
PR 17-JUN-1993; US-078683.
PA (CHIL-) CHILDRENS MEDICAL CENT.
PA (STRD) UNIV LELAND STANFORD JUNIOR.
PI Bernfield M, Kato M, Saunders S;
DR WPI; 95-052071/07.

```

PT DNA and protein sequences for recombinant syndecan-derived
 PT sites for heparin sulphate glycosaminoglycan side chains.
 PS Claim 15; Page 82; 97pp; English.
 CC Peptides R66797-802 correspond to varying length fragments of the
 CC N-terminal sequence of the mature mouse syndecan-1 (see R66793 and
 CC R66796). The sequence given here covers the amino acids from 1-154 of the
 CC mature protein sequence. The fragments are portions of the extracellular
 CC domain of the protein and contain a heparan sulphate glycosaminoglycan
 CC attachment site (R66795). The functional domains, esp. the soluble
 CC extracellular or heparan binding site, of the syndecan molecules (see
 CC R66797-812 and R66818) can be used to construct chimaeras by linking them
 CC to biological effector molecules, cell surface receptors, drugs,
 CC antibodies, diagnostic agents or components of microorganisms.
 SQ Sequence 154 AA;

Query Match 67.2%; Score 41; DB 13; Length 154;
 Best Local Similarity 55.6%; Pred. No. 2.71e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 74 ekpeegpvp 82
 I::: :|||
 QY 2 ENPDSSEPV 10

RESULT 6

ID R66802 standard; Peptide; 188 AA.
 AC R66802;
 DT 11-SEP-1995 (first entry)
 DE N-terminal fragment (a.a. 1-188) of mature mouse syndecan-1.
 KW Cell surface; proteoglycan; syndecan; mouse; mammary; epithelial cell;
 KW ectodomain; NMuMG; glycosylation; heparan sulphate; chondroitin sulphate;
 KW glycosaminoglycan; chimaera; chimaeric molecule; effector molecule;
 KW receptor; drug; antibody; diagnostic agent.
 OS Mus musculus.
 PN WO9500633-A.
 PD 05-JAN-1995.
 PF 17-JUN-1994; U06920.
 PR 17-JUN-1993; US-078683.
 PA (CHIL-) CHILDRENS MEDICAL CENT.
 PA (STRD) UNIV LELAND STANFORD JUNIOR.
 PI Bernfield M, Kato M, Saunders S;
 DR WPI; 95-052071/07.
 PT DNA and protein sequences for recombinant syndecan-derived
 PT proteoglycans - comprising a core protein having glycosylation
 PT sites for heparin sulphate glycosaminoglycan side chains.
 PS Claim 15; Page 82; 97pp; English.

CC Peptides R66797-802 correspond to varying length fragments of the
 CC N-terminal sequence of the mature mouse syndecan-1 (see R66793 and
 CC R66796). The sequence given here covers the amino acids from 1-188 of the
 CC mature protein sequence. The fragments are portions of the extracellular
 CC domain of the protein and contain a heparan sulphate glycosaminoglycan
 CC attachment site (R66795). The functional domains, esp. the soluble
 CC extracellular or heparan binding site, of the syndecan molecules (see
 CC R66797-812 and R66818) can be used to construct chimaeras by linking them
 CC to biological effector molecules, cell surface receptors, drugs,
 CC antibodies, diagnostic agents or components of microorganisms.
 SQ Sequence 188 AA;

Query Match 67.2%; Score 41; DB 13; Length 188;
 Best Local Similarity 55.6%; Pred. No. 2.71e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 74 ekpeegpvp 82
 I::: :|||
 QY 2 ENPDSSEPV 10

RESULT 7

ID R66804 standard; Peptide; 228 AA.
 AC R66804;
 DT 11-SEP-1995 (first entry)
 DE N-terminal fragment (a.a. 1-228) of mature mouse syndecan-1.

KW Cell surface; proteoglycan; syndecan; mouse; mammary; epithelial cell;
 KW ectodomain; NMuMG; glycosylation; heparan sulphate; chondroitin sulphate;
 KW glycosaminoglycan; chimaera; chimaeric molecule; effector molecule;
 KW receptor; drug; antibody; diagnostic agent.
 OS Mus musculus.
 PN WO9500633-A.
 PD 05-JAN-1995.
 PF 17-JUN-1994; U06920.
 PR 17-JUN-1993; US-078683.
 PA (CHIL-) CHILDRENS MEDICAL CENT.
 PA (STRD) UNIV LELAND STANFORD JUNIOR.
 PI Bernfield M, Kato M, Saunders S;
 DR WPI; 95-052071/07.
 PT DNA and protein sequences for recombinant syndecan-derived
 PT proteoglycans - comprising a core protein having glycosylation
 PT sites for heparin sulphate glycosaminoglycan side chains.
 PS Claim 22; Page 83; 97pp; English.
 CC Peptides R66803-810 correspond to the extracellular domains of the mouse
 CC syndecans 1-4 or fragments of these domains. The sequence given here
 CC represents amino acids 1-228, thus covering the entire extracellular
 CC domain, of the mature mouse syndecan-1 protein (see R66793 and R66796).
 CC The peptides presented all contain a heparan sulphate glycosaminoglycan
 CC attachment site (see R66794). The functional domains, esp. the soluble
 CC extracellular or heparan binding site, of the syndecan molecules (see
 CC R66797-812 and R66818) can be used to construct chimaeras by linking them
 CC to biological effector molecules, cell surface receptors, drugs,
 CC antibodies, diagnostic agents or components of microorganisms.
 SQ Sequence 228 AA;

Query Match 67.2%; Score 41; DB 13; Length 228;
 Best Local Similarity 55.6%; Pred. No. 2.71e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 74 ekpeegpvp 82
 I::: :|||
 QY 2 ENPDSSEPV 10

RESULT 8

ID R66796 standard; Protein; 289 AA.
 AC R66796;
 DT 11-SEP-1995 (first entry)
 DE Novel mouse proteoglycan syndecan-1 putative mature protein sequence.
 KW Cell surface; proteoglycan; syndecan; mouse; mammary; epithelial cell;
 KW ectodomain; NMuMG; glycosylation; heparan sulphate; chondroitin sulphate;
 KW chimaera; chimaeric molecule; effector molecule; receptor; drug;
 KW antibody; diagnostic agent.
 OS Mus musculus.
 PH Key Location/Qualifiers
 FT domain 1..230 /label= extracellular domain
 FT domain 231..26 /label= intracellular domain
 FT domain 231..255 /note= "heparan sulphate attachment site"
 FT domain 256..289 /label= transmembrane domain
 FT domain 289..289 /label= intracellular domain

PN WO9500633-A.
 PD 05-JAN-1995.
 PF 17-JUN-1994; U06920.
 PR 17-JUN-1993; US-078683.
 PA (CHIL-) CHILDRENS MEDICAL CENT.
 PA (STRD) UNIV LELAND STANFORD JUNIOR.
 PI Bernfield M, Kato M, Saunders S;
 DR WPI; 95-052071/07.

PT DNA and protein sequences for recombinant syndecan-derived
 PT proteoglycans - comprising a core protein having glycosylation
 PT sites for heparin sulphate glycosaminoglycan side chains.
 PS Claim 13; Page 81; 97pp; English.
 CC The sequence of the mature cell surface proteoglycan - syndecan-1.
 CC The corresponding gene (Q81748) was isolated from a mouse mammary
 CC epithelial cell cDNA library in lambda gt10 using a rabbit antisera against
 CC the ectodomain of NMuMG mouse mammary epithelial cell syndecan-1. The

CC complete sequence of the gene encodes a protein(R66793) of 30-35 kD
 CC comprising: (1) a hydrophilic N-terminal extracellular domain (residues
 CC 23-252); (2) a hydrophobic transmembrane domain (residues 253-277) and
 CC (3) a hydrophilic C-terminal intracellular domain (residues 278-331).
 CC The protein contains a protease susceptible cleavage sequence
 CC extracellularly and adjacent to the transmembrane region and at least one
 CC glycosylation site (residues 45-48) for attachment of a heparan sulphate
 CC chain to the extracellular region. The syndecan-1 protein is thought to
 CC contain a 22 amino acid signal peptide, lacking in this sequence, but
 CC this was unresolved due to N-terminal blocking of the mature peptide,
 CC however a likely site for signal peptidase cleavage occurs after Pro-22.
 CC The functional domains, esp. the soluble extracellular or heparan binding
 CC site, of the syndecan molecules (see R66797-812 and R66818) can be used
 CC to construct chimaeras by linking them to biological effector molecules,
 CC cell surface receptors, drugs, antibodies, diagnostic agents or
 CC components of microorganisms.
 CC Sequence 289 AA;

Query Match 67.2%; Score 41; DB 13; Length 289;
 Best Local Similarity 55.6%; Pred. No. 2.71e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 74 ekpeegp 82
 I::: :|||
 QY 2 ENPDSSEPV 10

RESULT 9
 ID W47156 standard; Protein; 311 AA.

AC W47156;
 DT 28-MAY-1998 (first entry)
 DE Mouse syndecan protein sequence.
 KW Syndecan; tumour suppression; tissue regeneration; enhancement;
 KW mouse; wound healing.

OS Mus sp.
 PN US5726058-A.
 PD 10-MAR-1998.
 PF 07-JUN-1995; 472217.
 PR 07-MAR-1994; US-206186.
 PR 01-DEC-1992; US-988427.
 PR 01-DEC-1993; WO-FI0514.
 PR 07-JUN-1995; US-472217.
 PA (ALAN)/ ALANEN-KURKI L.
 PA (AUVI)/ AUVINEN P.
 PA (JAAK)/ JAAKKOLA P.
 PA (JALK)/ JALKANEN M.
 PA (LEPP)/ LEPPAE S.
 PA (MALI)/ MALI M.
 PA (VIHI)/ VIHINEN T.
 PA (WAER)/ WAERRI A.
 PI Alanen-kurki L, Auvinen P, Jaakkola P, Jalkanen M,
 PI Leppae S, Mali M, Vihtinen T, Waerri A;
 DR WPI: 98-192770/17.
 DR N-PSDB: V15946.

PT New mouse syndecan gene sequences - useful for, e.g. suppressing
 PT tumour growth or promoting tissue regeneration in processes such as
 PT wound healing
 PS Disclosure; Fig 2A-O: 48pp; English.
 CC This is the mouse syndecan protein sequence. A 350 base pair fragment
 CC (V15948) of a purified 2196 base pair DNA molecule (V15947) enhances
 CC the expression of a gene operably linked to the promoter of the mouse
 CC syndecan gene in 3f3 cells following treatment with rGF- beta and bFGF
 CC when the fragment is operably linked to the promoter. A purified DNA
 CC molecule comprising a portion of the nucleotide residues 3538-3888 of
 CC the mouse syndecan genomic sequence suppresses expression of a gene
 CC operably linked to the promoter of the mouse syndecan gene in 3f3 cells
 CC treated with testosterone. Host cells can be transfected with vectors
 CC which contain either the enhancing or suppressing DNA molecules. The
 CC products may be used to alter the differentiated state of a host cell by
 CC altering its expression of syndecan, e.g. to induce and regulate syndecan
 CC expression, especially in cells which exhibit a malignant phenotype, or
 CC regardless of the origin of transformation. The products can be used to
 CC produce therapeutics for suppressing tumour growth. They may enhance the

CC syndecan expression in a host cell, by enhancing its gene transcription,
 CC especially in malignant or normal cells, and therefore promote tissue
 CC regeneration, especially in processes such as wound healing.
 SQ Sequence 311 AA;

Query Match 67.2%; Score 41; DB 29; Length 311;
 Best Local Similarity 55.6%; Pred. No. 2.71e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 96 ekpeegp 104
 I::: :|||
 QY 2 ENPDSSEPV 10

RESULT 10

ID R66793 standard; Protein; 311 AA.

AC R66793;
 DT 11-SEP-1995 (first entry)
 DE Novel mouse proteoglycan syndecan-1 protein sequence.
 KW Cell surface; proteoglycan; syndecan; mouse; mammary; epithelial cell;
 KW ectodomain; NMuMG; glycosylation; heparan sulphate; chondroitin sulphate;
 KW chimaera; chimaeric molecule; effector molecule; receptor; drug;
 KW antibody; diagnostic agent.
 OS Mus musculus.

FN Key Location/Qualifiers
 FT peptide 1..22
 FT /label= putative signal peptide
 FT domain 23..252
 FT /label= extracellular domain
 FT domain 45..48
 FT /label= heparan sulphate attachment site
 FT domain 253..277
 FT /label= transmembrane domain
 FT domain 278..311
 FT /label= intracellular domain

PN W09500633-A.

PD 05-JAN-1995.
 PF 17-JUN-1994; U06920.
 PR 17-JUN-1993; US-078683.
 PA (CHIL-) CHILDRENS MEDICAL CENT.
 PA (STRD) UNIV LELAND STANFORD JUNIOR.
 PI Bernfield M, Kato M, Saunders S;
 DR WPI: 95-052071/07.
 DR N-PSDB: Q81748.
 PT DNA and protein sequences for recombinant syndecan-derived
 PT proteoglycans - comprising a core protein having glycosylation
 PT sites for heparin sulphate glycosaminoglycan side chains.
 PS Claim 13; Page 80: 97pp; English.

CC The sequence of the novel cell surface proteoglycan - syndecan-1. The
 CC corresponding gene was isolated from a mouse mammary epithelial cell cDNA
 CC library in lambda gt11 using rabbit antisera against the ectodomain of
 CC NMuMG mouse mammary epithelial cell syndecan-1. The complete sequence of
 CC the gene encodes a protein of 30-35 kD comprising: (1) a hydrophilic
 CC N-terminal extracellular domain (residues 23-252); (2) a hydrophobic
 CC transmembrane domain (residues 253-277) and (3) a hydrophilic C-terminal
 CC intracellular domain (residues 278-331). The protein contains a protease
 CC susceptible cleavage sequence extracellularly and adjacent to the
 CC transmembrane region and at least one glycosylation site (residues 45-48)
 CC for attachment of a heparan sulphate chain to the extracellular region.
 CC The syndecan-1 protein is thought to contain a 22 amino acid signal
 CC peptide but this was unresolved due to N-terminal blocking of the mature
 CC peptide, however a likely site for signal peptidase cleavage occurs after
 CC Pro-22. The functional domains, esp. the soluble extracellular or
 CC heparan binding site, of the syndecan molecules (see R66797-812 and
 CC R66818) can be used to construct chimaeras by linking them to biological
 CC effector molecules, cell surface receptors, drugs, antibodies, diagnostic
 CC agents or components of microorganisms.
 SQ Sequence 311 AA;

Query Match 67.2%; Score 41; DB 13; Length 311;
 Best Local Similarity 55.6%; Pred. No. 2.71e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 96 ekpeesepv 104
QY 2 ENPDSSEPV 10

RESULT 11
ID R55276 standard; Protein; 311 AA.
AC R55276;
DT 08-DEC-1994 (first entry)
DE Syndecan protein.
KW Syndecan; enhancer; differentiation; tumor; therapy; hair growth.
OS Mus sp.
PN WO9412162-A.
PD 09-JUN-1994.
PF 01-DEC-1993; F10514.
PR 01-DEC-1992; US-988427.
PA (WAER/) WAER/ A M.
PA (ALAN/) ALANEN-KURKI L M.
PA (AUVI/) AUVINEN P O V.
PI Alanen-Kurki LM, Auvinen POV, Jaakkola PM, Jalkanen MT;
PI Leppaesm, Mali MS, Vihinen TA, Waerri AM;
DR WPI: 94-199926/24.
DR N-PSDB; Q67902.
PT Syndecan stimulation of cellular differentiation - useful for
PT decreasing tumour growth used to promote hair growth
PS Disclosure; Page 22-39; 65pp; English.
CC The mouse syndecan gene enhancer, located 8-10 kb upstream from the
CC initiation site, is given in Q67901. Manipulation of the enhancer
CC can be used either to slow or prevent tumor growth or to promote
CC differentiation of specific cell types, e.g. epidermal cells to
CC promote hair formation. The complete mouse syndecan gene and its
CC encoded protein are given in Q67902 and R55276.
SQ Sequence 311 AA;

Query Match 67.2%; Score 41; DB 10; Length 311;
Best Local Similarity 55.6%; Pred.No. 2.71e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 96 ekpeesepv 104
QY 2 ENPDSSEPV 10

RESULT 12
ID R87001 standard; Protein; 311 AA.
AC R87001;
DT 13-AUG-1996 (first entry)
DE Mouse syndecan-1.
KW Syndecan-1; ectodomain; tumour; cancer; therapy; proteoglycan;
KW cell differentiation.
OS Mus sp.
PN WO9534316-A1.
PD 21-DEC-1995.
PF 13-JUN-1995; F10344.
PR 13-JUN-1994; US-258862.
PA (JALK/) JALKANEN M.
PA (MALI/) MALI M.
PI Jalkanen M, Mali M;
DR WPI: 96-049416/05.
DR N-PSDB; T08126.
PT New syndecan ectodomain compsn. to reduce tumour growth - for
PT delivery to extracellular environment for suppressing tumour growth
PT in malignant and non-malignant tumours
PS Disclosure; Fig 2; 49pp; English.
CC Mouse syndecan-1 (R87001) is a cell surface proteoglycan. Its
CC ectodomain is capable of normalising the growth rate and
CC differentiation state of malignant cells. By providing the ectodomain
CC to a cell's extracellular environment, tumours characterised by a
CC loss of syndecan-1 can be treated. Expression of the ectodomain in
CC transfected S115 cells normalised malignant growth of the cells. The
CC ectodomain was also demonstrated to suppress tumour in vivo growth.
SQ Sequence 311 AA;

Query Match 67.2%; Score 41; DB 10; Length 311;
Best Local Similarity 55.6%; Pred.No. 2.71e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 96 ekpeesepv 104
QY 2 ENPDSSEPV 10

RESULT 12
ID R87001 standard; Protein; 311 AA.
AC R87001;
DT 13-AUG-1996 (first entry)
DE Mouse syndecan-1.
KW Syndecan-1; ectodomain; tumour; cancer; therapy; proteoglycan;
KW cell differentiation.
OS Mus sp.
PN WO9534316-A1.
PD 21-DEC-1995.
PF 13-JUN-1995; F10344.
PR 13-JUN-1994; US-258862.
PA (JALK/) JALKANEN M.
PA (MALI/) MALI M.
PI Jalkanen M, Mali M;
DR WPI: 96-049416/05.
DR N-PSDB; T08126.
PT New syndecan ectodomain compsn. to reduce tumour growth - for
PT delivery to extracellular environment for suppressing tumour growth
PT in malignant and non-malignant tumours
PS Disclosure; Fig 2; 49pp; English.
CC Mouse syndecan-1 (R87001) is a cell surface proteoglycan. Its
CC ectodomain is capable of normalising the growth rate and
CC differentiation state of malignant cells. By providing the ectodomain
CC to a cell's extracellular environment, tumours characterised by a
CC loss of syndecan-1 can be treated. Expression of the ectodomain in
CC transfected S115 cells normalised malignant growth of the cells. The
CC ectodomain was also demonstrated to suppress tumour in vivo growth.
SQ Sequence 311 AA;

Query Match 67.2%; Score 41; DB 17; Length 311;
Best Local Similarity 55.6%; Pred.No. 2.71e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 96 ekpeesepv 104
QY 2 ENPDSSEPV 10

RESULT 13
ID W18580 standard; Protein; 919 AA.
AC W18580;
DT 19-SEP-1997 (first entry)
DE Potato alpha-glucosidase.
KW Alpha-glucosidase; potato; starch; flavour; transgenic plant.
OS Solanum tuberosum cv. Record.
PN WO9724448-A1.
PD 10-JUL-1997.
PF 24-DEC-1996; G03239.
PR 28-DEC-1995; GB-026613.
PA (NICK-) NICKERSON BIOCHEM LTD.
PI Davies HV, Taylor MA;
DR WPI: 97-363683/33.
DR N-PSDB; T69925.

PT Sense and antisense plant alpha-glucosidase nucleic acids - useful
PT for modifying alpha-glucosidase activity, e.g. for enhancing plant
PT tissue flavour, producing novel starch, to control pre-sprouting,
PT etc.
PS Example 1; Fig 2; 40pp; English.
CC A 105.4 kDa alpha-glucosidase (W18580) is encoded by an isolated
CC nucleic acid (T69925) of potato cv. Record. This isolated nucleic
CC acid can be used to modify alpha-glucosidase activity in plants or
CC transformed microorganisms. Activity is enhanced through the use
CC of sense nucleic acids, thereby providing more efficient breakdown
CC of starch (useful in malting and brewing), novel starches, or
CC enhanced flavour to plant tissues such as fruit. Inhibition of
CC alpha-glucosidase activity through the use of antisense sequences
CC can inhibit starch breakdown and to control pre-sprouting and dry
CC weight/starch loss in cooking, storage and processing.
SQ Sequence 919 AA;

Query Match 67.2%; Score 41; DB 23; Length 919;
Best Local Similarity 55.6%; Pred.No. 2.71e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 70 eenpeseq 78
QY 1 QENPDSSEP 9

RESULT 14
ID W42736 standard; Peptide; 10 AA.
AC W42736;
DT 27-APR-1998 (first entry)
DE Antigenic decapeptide derived from the human Spl7 protein.
KW Spl7; sperm autoantigen; antigen; antigenic epitope; fertility;
KW immunocontraception; immunocontraceptive vaccine; autoimmune fertility.
OS Synthetic.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Modified_site 1 /note= "N-terminal acetyl"
FT WO9739020-A2.
PN 23-OCT-1997.
PD 11-APR-1997; U06489.
PF 15-APR-1996; US-632535.
PR (UYNC-) UNIV NORTH CAROLINA.
PA Lea I, ORand MG, Wildgren EE;
PI WPI: 97-526394/48.
DR Sperm protein Spl7 antigenic peptide(s) - used as
PT immunocontraceptive to reduce fertility
PS Example 14; Page 96; 160pp; English.
CC Peptides W42726-58 are a series of N-terminal acetylated sequential

CC decapeptides corresponding to fragments of the human Sp17 protein.
 CC Mimotope analysis was performed using the peptide pinblock method.
 CC Peptides comprising a contiguous segment of the amino acids of Sp17
 CC have been found to be antigenic. They induce antibodies which recognise
 CC sperm, and inhibit fertilisation. The peptides can be used in
 CC immunocontraceptive methods as a immunocontraceptive vaccine to reduce
 CC the fertility of animals, particularly in females. The presence of the
 CC antigenic peptides can be detected to diagnose autoimmune fertility in
 CC both male and female subjects.
 SQ Sequence 10 AA;

Query Match 65.6%; Score 40; DB 27; Length 10;
 Best Local Similarity 55.6%; Pred. No. 3.43e+02;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 2 qepekssdp 10
 || | : | : |
 QY 1 QENPDSSEP 9

RESULT 15
 ID W42824 standard; Peptide; 10 AA.
 AC W42824;
 DT 27-APR-1998 (first entry)
 DE Antigenic decapeptide derived from the human Sp17 protein.
 KW Sp17; sperm autoantigen; antigen; antigenic epitope; fertility;
 KW immunocontraception; immunocontraceptive vaccine; autoimmune fertility.
 OS Synthetic.
 OS Homo sapiens.
 PN W09739020-A2.
 PD 23-OCT-1997.
 PF 11-APR-1997; U06489.
 PR 15-APR-1996; US-632535.
 PA (UINC-) UNIV NORTH CAROLINA.
 PI Lea I, Orand MG, Widgren EE;
 DR WPI; 97-526394/48.
 PT Sperm protein Sp17 antigenic peptide(s) - used as
 PT immunocontraceptive to reduce fertility
 PS Example 25; Table 5; 160pp; English.
 CC Peptides W42774-867 are a series of sequential decapeptides corresponding
 CC to fragments of the human Sp17 protein. Each decapeptide is shifted along
 CC the Sp17 sequence by one amino acid. W42800-820 are derived from the
 CC human Sp17 sequence between amino acids 40-80 inclusive, while W42849-67
 CC are between amino acids 114-149 inclusive. The peptides contain antigenic
 CC epitopes in mice, rabbits, non-human primates and humans. They induce
 CC antibodies which recognise sperm, and inhibit fertilisation. The peptides
 CC can be used in immunocontraceptive methods as a immunocontraceptive
 CC vaccine to reduce the fertility of animals, particularly in females. The
 CC presence of the antigenic peptides can be detected to diagnose autoimmune
 CC fertility in both male and female subjects.
 SQ Sequence 10 AA;

Query Match 65.6%; Score 40; DB 28; Length 10;
 Best Local Similarity 55.6%; Pred. No. 3.43e+02;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1 qepekssdp 9
 || | : | : |
 QY 1 QENPDSSEP 9

Search completed: Thu May 13 15:44:32 1999
 Job time : 18 secs.

Result No.	Score	Query %		Length	DB	ID	Description	Pred. No.
		Match	Ident					
1	41	100.0	302	5	O44687	T22D1.2 PROTEIN.	1.45e+00	
2	38	92.7	357	13	P73765	NEUROD-1 LIKE PROTEIN.	9.35e+00	
3	38	92.7	357	13	O73826	NEUROD-1 LIKE PROTEIN.	9.35e+00	
4	38	92.7	539	2	Q54398	PROTEINASE.	9.35e+00	
5	38	92.7	1404	4	O15080	KIAA0375.	9.35e+00	
6	37	90.2	102	1	O07119	PUTATIVE TRYPTOPHANYL	1.70e+01	
7	37	90.2	329	14	Q98824	STRAIN 13881V VP6 (S3)	1.70e+01	
8	37	90.2	329	14	Q96652	STRAIN 11C81Z VP6 (S3)	1.70e+01	
9	37	90.2	329	14	Q96659	STRAIN 11C81Z VP6 (S3)	1.70e+01	
10	37	90.2	329	14	Q96658	STRAIN BTV 13 VP6 (S3)	1.70e+01	
11	37	90.2	329	14	Q96661	STRAIN BTV 13 VP6 (S3)	1.70e+01	
12	37	90.2	329	14	Q96648	VACCINE STRAIN VP6 (S3)	1.70e+01	
13	37	90.2	329	14	Q96656	STRAIN 10B80Y VP6 (S3)	1.70e+01	
14	37	90.2	329	14	Q96657	STRAIN 17C81W VP6 (S3)	1.70e+01	
15	37	90.2	329	14	Q96657	STRAIN BTV 11 VP6 (S3)	1.70e+01	
16	37	90.2	329	14	Q96654	STRAIN 11UC8 VP6 (S3)	1.70e+01	
17	37	90.2	329	14	Q96653	STRAIN 11UC2 VP6 (S3)	1.70e+01	
18	37	90.2	329	14	Q96655	STRAIN 13B89Y VP6 (S3)	1.70e+01	
19	37	90.2	329	14	Q96651	STRAIN 11081Z VP6 (S3)	1.70e+01	
20	37	90.2	329	14	Q96649	STRAIN 10080V VP6 (S3)	1.70e+01	
			329	14	Q96649	STRAIN 10B90Z VP6 (S3)	1.70e+01	

QY 3 PTGEPQ 8

RESULT 2

ID P79765 PRELIMINARY; PRT; 357 AA.

AC P79765;

DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
 DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE NEUROD-LIKE PROTEIN.

OS GALLUS GALLUS (CHICKEN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
 OC GALLIFORMES.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-WHITE LEHORN; TISSUE-RETINA;

RA ROTOCIL T., MATTER-SADZINSKI L., ALLIOD C., BALLIVET M., MATTER J.M.;

RL DEVELOPMENT 124:3263-3272(1997).

DR EMBL; Y09596; E283389; -.

DR PFAM; PF00010; HLH.

SQ SEQUENCE 357 AA; 38809 MW; 04A96823 CRC32;

Query Match

Best Local Similarity 92.7%; Score 38; DB 13; Length 357;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 10 PAGEPQ 15

I:||||

QY 3 PTGEPQ 8

RESULT 3

ID O73826 PRELIMINARY; PRT; 357 AA.

AC O73826;

DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE NEUROD.

OS GALLUS GALLUS (CHICKEN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
 OC GALLIFORMES.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-WHITE LEHORN;

RA YAN R.-T., WANG S.-Z.;

RL J. NEUROBIOL. 0:0-0(1998).

DR EMBL; AF060885; G3094020; -.

SQ SEQUENCE 357 AA; 38858 MW; 7D406FF3 CRC32;

Query Match

Best Local Similarity 92.7%; Score 38; DB 13; Length 357;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 10 PAGEPQ 15

I:||||

QY 3 PTGEPQ 8

RESULT 4

ID Q54398 PRELIMINARY; PRT; 539 AA.

AC Q54398;

DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE PROTEINASE.

OS SLPD.

OC STREPTOMYCES LIVIDANS.

OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-66;

RX MEDLINE; 96042075.

RA BINNIE C., BUTLER M.J., APHALE J.S., BOURGAULT R., DIZONNO M.A.,
 RA KRYGSMAN P., LIAO L., WALCZYK E., MALEK L.T.,
 RL J. BACTERIOL. 177:6033-6040(1995).

DR EMBL; L42758; G940303; -.

DR PFAM; PF00561; abhydrolase.

SQ SEQUENCE 539 AA; 57033 MW; 6EAA9AC1 CRC32;

Query Match

Best Local Similarity 92.7%; Score 38; DB 2; Length 539;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 457 PTGEPH 462

I:||||

QY 3 PTGEPQ 8

RESULT 5

ID O15080 PRELIMINARY; PRT; 1404 AA.

AC O15080;

DT 01-JAN-1998 (TREMBLREL. 05, CREATED)

DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)

DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)

DE KIAA0375.

GN KIAA0375.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;

OC EUTHERIA; PRIMATES.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-BRAIN;

RX MEDLINE; 97349984.

RA NAGASE T., ISHIKAWA K., NAKAJIMA D., OHIRA M., SEKI N., MIYAJIMA N.,

RA TANAKA A., KOTANI H., NOMURA N., OHARA O.;

RL DNA RES. 4:141-150(1997).

DR EMBL; AB002373; D1021671; -.

SQ SEQUENCE 1404 AA; 149264 MW; 675DE431 CRC32;

Query Match

Best Local Similarity 92.7%; Score 38; DB 4; Length 1404;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 746 PSGEQ 751

I:||||

QY 3 PTGEPQ 8

RESULT 6

ID O07119 PRELIMINARY; PRT; 102 AA.

AC O07119;

DT 01-JUL-1997 (TREMBLREL. 04, CREATED)

DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)

DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

DE PUTATIVE TRYPTOPHYLYL AMINOACYL TRNA SYNTHETASE (FRAGMENT).

GN TRPS.

OS HALOBACTERIUM VOLCANII (HALOFERAX VOLCANII).

OC ARCHAEABACTERIA; EURYARCHAEOTA; HALOBACTERIALES; HALOBACTERIACEAE.

RN [1]

RP SEQUENCE FROM N.A.

RA KLEMAN-LEYER K., ARMSTRUSTER D.W., DANIELS C.J.;

RL CELL 89:839-847(1997).

DR EMBL; AF001578; G2160793; -.

KW AMINOACYL-TRNA SYNTHASE.

FT NON_TER 102 102

SQ SEQUENCE 102 AA; 11442 MW; E13C318F CRC32;

Query Match

Best Local Similarity 90.2%; Score 37; DB 1; Length 102;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 94 PTGDPH 99

I:||||

QY 3 PTGEPQ 8

RC	STRAIN-BTV 17;
RX	MEDLINE; 96357114.
RA	DE MATTOS C.C., DE MATTOS C.A., MACLACHLAN N.J., GIAVEDONI L.D.,
RY	YILMA T., OSBURN B.I.;
RL	J. VIROL. 70:5735-5739(1996).
DR	EMBL; U55798; G1497973; -.
SQ	SEQUENCE 329 AA; 35659 MW; 97FELB22 CRC32;
 Query Match 90.2%; Score 37; DB 14; Length 329; Best Local Similarity 66.7%; Pred. No. 1.70e+01; Matches 4; Conservative 2; Mismatches 0; Indels 0;	
Dd	289 PTGDPH 294 :::
Oy	3 PTGEPO 8 :::
 RESULT 10 ID Q96658 PRELIMINARY; PRT; 329 AA. AC Q96658; DT 01-FEB-1997 (TREMBREL. 02, CREATED) DT 01-FEB-1997 (TREMBREL. 02, LAST SEQUENCE UPDATE) DT 01-JUN-1998 (TREMBREL. 06, LAST ANNOTATION UPDATE) DE STRAIN BTV 13 VP6 (S3). DN S3. OS BLUETONGUE VIRUS. OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES. [1] RN RP SEQUENCE FROM N.A. RC STRAIN-BTV 13; RX MEDLINE; 96357114. RA DE MATTOS C.C., DE MATTOS C.A., MACLACHLAN N.J., GIAVEDONI L.D., RY YILMA T., OSBURN B.I.;	
RL	J. VIROL. 70:5735-5739(1996).
DR	EMBL; U55797; G1497971; -.
SQ	SEQUENCE 329 AA; 35542 MW; B581CF73 CRC32;
 Query Match 90.2%; Score 37; DB 14; Length 329; Best Local Similarity 66.7%; Pred. No. 1.70e+01; Matches 4; Conservative 2; Mismatches 0; Indels 0;	
Dd	289 PTGDPH 294 :::
Oy	3 PTGEPO 8 :::
 RESULT 11 ID Q96661 PRELIMINARY; PRT; 329 AA. AC Q96661; DT 01-FEB-1997 (TREMBREL. 02, CREATED) DT 01-FEB-1997 (TREMBREL. 02, LAST SEQUENCE UPDATE) DT 01-JUN-1998 (TREMBREL. 06, LAST ANNOTATION UPDATE) DE VACCINE STRAIN VP6 (S3). DN S3. OS BLUETONGUE VIRUS. OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES. [1] RN RP SEQUENCE FROM N.A. RC STRAIN-VACCINE STRAIN; RX MEDLINE; 96357114. RA DE MATTOS C.C., DE MATTOS C.A., MACLACHLAN N.J., GIAVEDONI L.D., RY YILMA T., OSBURN B.I.;	
RL	J. VIROL. 70:5735-5739(1996).
DR	EMBL; U55800; G1497977; -.
SQ	SEQUENCE 329 AA; 35517 MW; 651408E6 CRC32;
 Query Match 90.2%; Score 37; DB 14; Length 329; Best Local Similarity 66.7%; Pred. No. 1.70e+01; Matches 4; Conservative 2; Mismatches 0; Indels 0;	
Dd	289 PTGDPH 294 :::
Ov	3 PTGEPO 8 :::

```
RESULT 12
ID Q96648 PRELIMINARY; PRT; 329 AA.
AC Q96648;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE STRAIN 1080Y VP6 (S3).
GN S3.
OS BLUETONGUE VIRUS.
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-1080Y;
RX MEDLINE; 96357114.
RA DE MATTOS C.C., DE MATTOS C.A., MACLACHLAN N.J., GIAVEDONI L.D.,
RA YILMA T., OSBURN B.I.;
RL J. VIROL. 70:5735-5739(1996).
DR EMBL; U55778; G1497933; -.
SQ SEQUENCE 329 AA; 35585 MW; 95C25B6C CRC32;

Query Match 90.2%; Score 37; DB 14; Length 329;
Best Local Similarity 66.7%; Pred. No. 1.70e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294
QY 3 PTGEPQ 8

RESULT 13
ID Q96656 PRELIMINARY; PRT; 329 AA.
AC Q96656;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE STRAIN 17C81W VP6 (S3).
GN S3.
OS BLUETONGUE VIRUS.
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-17C81W;
RX MEDLINE; 96357114.
RA DE MATTOS C.C., DE MATTOS C.A., MACLACHLAN N.J., GIAVEDONI L.D.,
RA YILMA T., OSBURN B.I.;
RL J. VIROL. 70:5735-5739(1996).
DR EMBL; U55793; G1497963; -.
SQ SEQUENCE 329 AA; 35535 MW; 5C96CCD9 CRC32;

Query Match 90.2%; Score 37; DB 14; Length 329;
Best Local Similarity 66.7%; Pred. No. 1.70e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294
QY 3 PTGEPQ 8

RESULT 14
ID Q96657 PRELIMINARY; PRT; 329 AA.
AC Q96657;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE STRAIN BTV 11 VP6 (S3).
GN S3.
OS BLUETONGUE VIRUS.
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BTV 11;
```

```
RX MEDLINE; 96357114.
RA DE MATTOS C.C., DE MATTOS C.A., MACLACHLAN N.J., GIAVEDONI L.D.,
RA YILMA T., OSBURN B.I.;
RL J. VIROL. 70:5735-5739(1996).
DR EMBL; U55796; G1497969; -.
SQ SEQUENCE 329 AA; 35629 MW; A645E255 CRC32;

Query Match 90.2%; Score 37; DB 14; Length 329;
Best Local Similarity 66.7%; Pred. No. 1.70e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294
QY 3 PTGEPQ 8

RESULT 15
ID Q96654 PRELIMINARY; PRT; 329 AA.
AC Q96654;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE STRAIN 11UC8 VP6 (S3).
GN S3.
OS BLUETONGUE VIRUS.
OC VIRIDAE; DS-RNA NONENVELOPED VIRUSES; REOVIRIDAE; ORBIVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-11UC8;
RX MEDLINE; 96357114.
RA DE MATTOS C.C., DE MATTOS C.A., MACLACHLAN N.J., GIAVEDONI L.D.,
RA YILMA T., OSBURN B.I.;
RL J. VIROL. 70:5735-5739(1996).
DR EMBL; U55787; G1497951; -.
SQ SEQUENCE 329 AA; 35573 MW; 1B421073 CRC32;

Query Match 90.2%; Score 37; DB 14; Length 329;
Best Local Similarity 66.7%; Pred. No. 1.70e+01;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 289 PTGDPH 294
QY 3 PTGEPQ 8

Search completed: Thu May 13 15:42:06 1999
Job time : 34 secs.
```

[W][O][R][D][S] (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 13 15:28:47 1999; MasPar time 2.03 Seconds
Tabular output not generated. 132.450 Million cell updates/sec

Title: >US-09-040-485-6
Description: (1-10) from US09040485.pep
Perfect Score: 64
Sequence: 1 APPEDNPVED 10

Scoring table: PAM 150
Gap 15

Searched: 74019 seqs, 26840295 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot35
1:swissprot

Statistics: Mean 21.884; Variance 24.812; scale 0.882

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match %	Length DB	ID Description Pred. No.
1	64	100.0	757	1 ASPH_HUMAN ASPARTYL/ASPARAGINYL B 2.44e-04
2	50	78.1	704	1 KPC1_CAEEL PROTEIN KINASE C-LIKE 7.01e-01
3	46	71.9	562	1 LCB2_KLULA SERINE PALMITOYLTRANSF 5.49e+00
4	45	70.3	365	1 TRBP_MOUSE PROTAMINE-1 RNA BINDIN 9.01e+00
5	45	70.3	928	1 RB_HUMAN RETINOBLASTOMA-ASSOCIA 9.01e+00
6	45	70.3	932	1 SYL_ARCFU LEUCYL-TRNA SYNTHETASE 9.01e+00
7	45	70.3	2839	1 NFI_HUMAN NEUROFIBROMIN (NEUROFI 9.01e+00
8	45	70.3	2841	1 NFI_MOUSE NEUROFIBROMIN (NEUROFI 9.01e+00
9	44	68.8	36	1 PAHO_LARAR PANCREATIC HORMONE (PA 1.47e+01
10	44	68.8	36	1 PAHO_STRCA PANCREATIC HORMONE (PA 1.47e+01
11	44	68.8	80	1 PAHO_CHICK PANCREATIC HORMONE PRE 1.47e+01
12	44	68.8	122	1 YLC3_YEREN HYPOTHETICAL 13.7 KD P 1.47e+01
13	44	68.8	213	1 NEUM_CARAU NEUROMODULIN (AXONAL M 1.47e+01
14	44	68.8	262	1 TRT3_CHICK TROPONIN T, FAST SKELE 1.47e+01
15	44	68.8	376	1 ACTM_APLCA ACTIN, MUSCLE 1.47e+01
16	44	68.8	395	1 YACE_VIBCH HYPOTHETICAL 44.0 KD P 1.47e+01
17	44	68.8	461	1 YAU2_SCHPO HYPOTHETICAL 52.9 KD P 1.47e+01
18	44	68.8	647	1 DPOG_CHICK DNA POLYMERASE GAMMA (1.47e+01
19	44	68.8	979	1 RFX1_HUMAN MHC CLASS II REGULATOR 1.47e+01
20	43	67.2	307	1 OOX2_ACEAC UBIQUINOL OXIDASE POLY 2.37e+01
21	43	67.2	491	1 NWT_CRYNE GLYCYLPEPTIDE N-TETRA 2.37e+01
22	43	67.2	506	1 GAG_SIVMK GAG POLYPROTEIN (CORE 2.37e+01
23	43	67.2	507	1 GAG_SIVSP GAG POLYPROTEIN (CORE 2.37e+01

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 281 APPENDPVED 290
QY 1 APPENDPVED 10

RESULT 2
ID KPC1_CAEEL STANDARD; PRT; 704 AA.
AC P34722;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PROTEIN KINASE C-LIKE 1 (EC 2.7.1.-) (PKC).
GN TPA-1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 95387388.
RA SANO T., TABUSE Y., NISHIWAKI K., MIWA J.;
RL J. MOL. BIOL. 251:477-485(1995).
RN [2]
RP SEQUENCE OF 148-704 FROM N.A.
RX MEDLINE; 89186920.
RA TABUSE Y., NISHIWAKI K., MIWA J.;
RL SCIENCE 243:1713-1716(1989).
RN [3]
RP REVISIONS TO REF. 2.
RA MIWA J.;

RL SUBMITTED (MAR-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: PKC IS ACTIVATED BY DIACYLGLYCEROL WHICH IN TURN
CC PHOSPHORYLATES A RANGE OF CELLULAR PROTEINS. PKC ALSO SERVES AS
CC THE RECEPTOR FOR PHORBOL ESTERS, A CLASS OF TUMOR PROMOTERS.
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS (TPA-1A AND TPA-1B) ARE PRODUCED
CC BY ALTERNATIVE SPLICING OF THE SAME GENE. THE SEQUENCE SHOWN IS
CC THAT OF TPA-1A.
CC -!- SIMILARITY: CONTAINS TWO COPIES OF THE ZINC-DEPENDENT PHORBOL-
CC ESTER AND DAG BINDING DOMAIN.
DR EMBL; D49525; G1217584; -.
DR EMBL; D49525; G1217585; -.
DR EMBL; D14815; G303529; -.
DR HSP; P05132; 1ATP.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00479; DAG_PE_BINDING_DOMAIN; 2.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW TRANSFERASE: SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING; ZINC;
KW PHORBOL-ESTER BINDING; REPEAT; ALTERNATIVE SPLICING; PHOSPHORYLATION.
FT DOMAIN 166 215
FT DOMAIN 238 287
FT DOMAIN 375 634
FT NP_BIND 381 389
FT BINDING 404 404
FT ACT_SITE 499 499
FT MOD_RES 89 89
FT MOD_RES 139 139
FT MOD_RES 324 324
FT VARSPPLIC 1 137
SQ SEQUENCE 704 AA; 80298 MW; 84720228 CRC32;

Query Match 78.1%; Score 50; DB 1; Length 704;
Best Local Similarity 70.0%; Pred. No. 7.01e-01;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 127 APDDHPVED 136
QY 1 APPENDPVED 10

RESULT 3
ID LCB2_KLULA STANDARD; PRT; 562 AA.

AC P48241;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE SERINE PALMITOYLTRANSFERASE 2 (EC 2.3.1.50) (LONG CHAIN BASE
DE BIOSYNTHESIS PROTEIN 2) (SPT 2).
GN LCB2.
OS KLUYVEROMYCES LACTIS (YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAG;
RX MEDLINE; 97080528
RA NAGIEC M.M., LESTER R.L., DICKSON R.C.;
RL GENE 177:237-241(1996).
CC -!- CATALYTIC ACTIVITY: PALMITOYL-COA + L-SERINE = COA + 3-DEHYDRO-
CC D-SPHINGANINE + CO(2).
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -!- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF THE LONG-CHAIN BASE
CC COMPONENT OF SPHINGOLIPIDS.
CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
CC -!- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.
DR EMBL; U15646; G1001949; -.
DR PROSITE; PS00599; AA_TRANSFER_CLASS_2; 1.
KW TRANSFERASE; ACYLTRANSFERASE; TRANSMEMBRANE; PYRIDOXAL PHOSPHATE.
FT TRANSMEM 61 81
FT TRANSMEM 356 376
FT BINDING 365 365
SQ SEQUENCE 562 AA; 63004 MW; 901ACF38 CRC32;

Query Match 71.9%; Score 46; DB 1; Length 562;
Best Local Similarity 77.8%; Pred. No. 5.49e+00;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 13 PREDIPLED 21
QY 2 PPDNPVED 10

RESULT 4
ID TRBP_MOUSE STANDARD; PRT; 365 AA.
AC P97473;

DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROTAMINE-1 RNA BINDING PROTEIN (PRM-1 RNA BINDING PROTEIN).
GN TARBP2 OR PRBP.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS.
RX MEDLINE; 96220490.
RA LEE K., FAJARDO M.A., BRAUN R.E.;
RL MOL. CELL. BIOL. 16:3023-3034(1996).
CC -!- FUNCTION: BINDS IN VITRO TO THE PRM1 3'UTR. SEEMS TO ACT AS A
CC GENERAL REPRESSOR OF TRANSLATION.
CC -!- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC.
CC -!- SIMILARITY: CONTAINS 3 DRBM (DOUBLE-STRANDED RNA-BINDING) DOMAINS.
DR EMBL; U79962; G1737216; -.
DR MGD; MGI:108201; PRBP.
KW RNA-BINDING; REPEAT; NUCLEAR PROTEIN.
FT DOMAIN 79 95
FT DOMAIN 208 224
FT DOMAIN 342 358
SQ SEQUENCE 365 AA; 38789 MW; 0AF5653D CRC32;

Query Match 70.3%; Score 45; DB 1; Length 365;
Best Local Similarity 75.0%; Pred. No. 9.01e+00;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 117 SPEDTPV 124
:|||||
Qy 1 APPENDPV 8

RESULT 5
ID RE_HUMAN STANDARD; PRT; 928 AA.
AC P06400; P78499;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE RETINOBLASTOMA-ASSOCIATED PROTEIN (P110) (P105-RB) (RB).
GN REL.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88014238.
RA LEE W.-H., SHEW J.-Y., HONG F.D., SERY T.W., DONOSO L.A., YOUNG L.-J.,
RA BOOKSTEIN R., LEE E.Y.-H.P.;
RL NATURE 329:642-645(1987).
RN [2]
RP REVISIONS.
RX MEDLINE; 87149066.
RA LEE W.-H., BOOKSTEIN R., HONG F.D., YOUNG L.-J., SHEW J.-Y.,
RA LEE E.Y.-H.P.;
RL SCIENCE 235:1394-1399(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88097427.
RA FRIEND S.H., HOROWITZ J.M., GERBER M.R., WANG X.-F., BOGENMANN E.,
RA LI F.P., WEINBERG R.A.;
RL PROC. NATL. ACAD. SCI. U.S.A. 84:9059-9063(1987).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90006771.
RA MCGEE T.L., YANDELL D.W., DRYJA T.P.;
RL GENE 80:119-128(1989).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94063891.
RA TOGUCHIDA J., MCGEE T.L., CICCARELLI J.A., EAGLE J.R., YANDELL D.W.,
RA DRYJA T.P.;
RL GENOMICS 17:535-543(1993).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE-CARCINOMA;
RX MEDLINE; 92319557.
RA HOGG A., ONADIM Z., BAIRD P.N., COWELL J.K.;
RL ONCOGENE 7:1445-1451(1992).
RN [7]
RP SEQUENCE OF 1-45 FROM N.A.
RX MEDLINE; 89239464.
RA T'ANG A., WU K.J., HASHIMOTO T., LIU W.Y., TAKAHASHI R., SHI X.H.,
RA M'HARA K., ZHANG F.H., CHEN Y.Y., DU C., QIAN J., LIN Y.G.,
RA MURPHEE A.L., QIU W.R., THOMPSON T., BENEDICT W.F., FUNG Y.K.T.;
RL ONCOGENE 4:401-407(1989).
RN [8]
RP PHOSPHORYLATION SITES.
RX MEDLINE; 92097548.
RA LEES J.A., BUCHKOVICH K.J., MARSHAK D.R., ANDERSON C.W., HARLOW E.;
RL EMBO J. 10:4279-4290(1991).
RN [9]
RP VARIANT LEU-567.
RX MEDLINE; 90081757.
RA YANDELL D.W., CAMPBELL T.A., DAYTON S.H., PETERSEN R., WALTON D.,
RA LITTLE J.B., MCCONKIE-ROSELL A., BUCKLEY E., DRYJA T.;
RL NEW ENGL. J. MED. 321:1689-1695(1989).
RN [10]
RP VARIANT TRP-661.
RX MEDLINE; 92335261.
RA ONADIM Z., HOGG A., BAIRD P.N., COWELL J.K.;

PROC. NATL. ACAD. SCI. U.S.A. 89:6177-6181(1992).
[11]
RN VARIANTS ASN-480 DEL AND TRP-661.
RX MEDLINE; 95012220.
RA LOHMANN D.R., BRANDT B., HOEPFING W., PASSARGE E., HORSTHEMKE B.;
RL HUM. GENET. 94:349-354(1994).
RN [12]
RP VARIANTS RB ASP-137 AND TRP-661.
RX MEDLINE; 97456418.
RA LOHMANN D.R., GERICK M., BRANDT B., OELSCHLAEGER U., LORENZ B.,
RA PASSARGE E., HORSTHEMKE B.;
RL AM. J. HUM. GENET. 61:282-294(1997).
CC -!- FUNCTION: PROBABLY ACTS AS A REGULATOR OF OTHER GENES. FORMS A
CC COMPLEX WITH ADENOVIRUS E1A AND WITH SV40 LARGE T ANTIGEN. ACTS AS
CC A TUMOR SUPPRESSOR. MAY BIND AND MODULATE FUNCTIONALLY CERTAIN
CC CELLULAR PROTEINS WITH WHICH T AND E1A COMPETE FOR POCKET BINDING.
CC POTENT INHIBITOR OF E2F-MEDIATED TRANS-ACTIVATION, INTERACTS
CC PREFERENTIALLY WITH TRANSCRIPTION FACTOR E2F1.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- PTM: PHOSPHORYLATED FROM S TO M PHASE OF THE CELL CYCLE AND IS
CC DEPHOSPHORYLATED IN G1. T, BUT NOT E1A, BINDS ONLY TO THE
CC UNPHOSPHORYLATED FORM.
CC -!- DISEASE: DELETION OR ALTERATION OF THIS PROTEIN RESULTS IN THE
CC CHILDHOOD CANCER RETINOBLASTOMA (RB). RB IS A CONGENITAL
CC MALIGNANT TUMOR THAT ARISES FROM THE NUCLEAR LAYERS OF THE
CC RETINA. IT OCCURS IN ABOUT 1:20 000 LIVE BIRTHS AND REPRESENTS
CC ABOUT 2% OF CHILDHOOD MALIGNANCIES. IT IS BILATERAL IN ABOUT 30%
CC OF CASES. ALTHOUGH MOST RB APPEAR SPORADICALLY, ABOUT 20% ARE
CC TRANSMITTED AS AN AUTOSOMAL DOMINANT TRAIT WITH INCOMPLETE
CC PENETRANCE. THE DIAGNOSIS IS USUALLY MADE BEFORE THE AGE OF 2
CC YEARS WHEN STRABISMUS OR A GREY TO YELLOW REFLEX FROM PUPIL ("CAT
CC EYE") IS INVESTIGATED.
CC -!- SIMILARITY: BELONGS TO THE RETINOBLASTOMA PROTEIN (RB) FAMILY.
CC EMBL; L41870; G793995; -
DR EMBL; M27866; G521212; -
DR EMBL; M27845; G521212; JOINED.
DR EMBL; M27846; G521212; JOINED.
DR EMBL; M27847; G521212; JOINED.
DR EMBL; M27849; G521212; JOINED.
DR EMBL; M27850; G521212; JOINED.
DR EMBL; M27851; G521212; JOINED.
DR EMBL; L35146; G521212; JOINED.
DR EMBL; M27852; G521212; JOINED.
DR EMBL; M27853; G521212; JOINED.
DR EMBL; M27854; G521212; JOINED.
DR EMBL; M27855; G521212; JOINED.
DR EMBL; M27856; G521212; JOINED.
DR EMBL; M27857; G521212; JOINED.
DR EMBL; M27858; G521212; JOINED.
DR EMBL; M27859; G521212; JOINED.
DR EMBL; M27860; G521212; JOINED.
DR EMBL; L35147; G521212; JOINED.
DR EMBL; M27862; G521212; JOINED.
DR EMBL; M27863; G521212; JOINED.
DR EMBL; M27864; G521212; JOINED.
DR EMBL; M27865; G521212; JOINED.
DR EMBL; M15400; G190959; JOINED.
DR EMBL; M28419; G190963; -
DR EMBL; M33647; G190946; -
DR EMBL; X16439; G35895; -
DR EMBL; L41997; G804760; -
DR EMBL; L41890; G793951; -
DR EMBL; L41891; G793953; -
DR EMBL; L41893; G793959; -
DR EMBL; L41894; G793961; -
DR EMBL; L41895; G793963; -
DR EMBL; L41896; G793965; -
DR EMBL; L41897; G793967; -
DR EMBL; L41898; G793971; -
DR EMBL; L41899; G793976; -
DR EMBL; L11910; G292421; -
DR PIR; JS0276; RBH0.
DR PIR; A39947; A39947.

DR TRANSFAC; T00722; -
 DR MIM; 180200; -
 DR MIM; 259200; -
 DR MIM; 259500; -
 KW TRANSCRIPTION REGULATION; DNA-BINDING; NUCLEAR PROTEIN;
 KW PHOSPHORYLATION; ANTI-ONCOGENE; DISEASE MUTATION.
 FT DOMAIN 10 18 POLY-ALA.
 FT DOMAIN 20 29 POLY-PRO.
 FT DOMAIN 373 771 POCKET (BINDS T AND EIA).
 FT DOMAIN 373 579 DOMAIN A.
 FT DOMAIN 580 639 SPACER.
 FT DOMAIN 640 771 DOMAIN B.
 FT MOD_RES 249 249 PHOSPHORYLATION (BY CDC2).
 FT MOD_RES 252 252 PHOSPHORYLATION (BY CDC2).
 FT MOD_RES 373 373 PHOSPHORYLATION (BY CDC2).
 FT MOD_RES 807 807 PHOSPHORYLATION (BY CDC2).
 FT MOD_RES 811 811 PHOSPHORYLATION (BY CDC2).
 FT VARIANT 137 137 E -> D (IN UNILATERAL RB).
 FT VARIANT 480 480 MISSING (IN RB; MILD FORM).
 FT VARIANT 567 567 S -> L (IN RB).
 FT VARIANT 661 661 R -> W (IN RB; MILD FORM).
 SQ SEQUENCE 928 AA; 106159 MW; 0BD3B72E CRC32;

Query Match 70.3%; Score 45; DB 1; Length 928;
 Best Local Similarity 50.0%; Pred. No. 9.01e+00;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 27 PPPEEDPEQD 36
 QY 1 APPEDNPVED 10

RESULT 6
 ID SYL_ARCFU STANDARD; PRT; 932 AA.
 AC O30250;
 DT 15-JUL-1998 (REL. 36, CREATED)
 DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE LEUCYL-TRNA SYNTHETASE (EC 6.1.1.4) (LEUCINE-TRNA LIGASE) (LEURS).
 GN LEUS OR AF2421.
 OS ARCHAEOGLOBUS FULGIDUS.
 OC ARCHAEBACTERIA; EURYARCHAEOTA; ARCHAEOGLOBALES; ARCHAEOGLOBACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE; 98049343.
 RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
 RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
 RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KIRPIDES N.C.,
 RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
 RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
 RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
 RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,
 RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SKES S.M.,
 RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
 RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,
 RA VENTER J.C.;
 RL NATURE 390:364-370(1997).
 CC -!- CATALYTIC ACTIVITY: ATP + L-LEUCINE + TRNA(LEU) = AMP +
 CC -!- PYROPHOSPHATE + L-LEUCYL-TRNA(LEU).
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
 DR EMBL; AE001108; G2650668; -;
 DR TIGR; AF2421; -;
 DR PROSITE; PS00178; AA_TRNA_LIGASE_1;
 KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
 FT SIMILAR 38 48 "HIGH" REGION.
 FT SIMILAR 630 634 "KMSKS" REGION.
 FT BINDING 633 633 ATP (BY SIMILARITY).
 SQ SEQUENCE 932 AA; 108628 MW; 13803D54 CRC32;

Query Match 70.3%; Score 45; DB 1; Length 932;
 Best Local Similarity 66.7%; Pred. No. 9.01e+00;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 180 PHQNPVED 188
 QY 2 PPEDNPVED 10
 RESULT 7
 ID NFL_HUMAN STANDARD; PRT; 2839 AA.
 AC P21359;
 DT 01-MAY-1991 (REL. 18, CREATED)
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE NEUROFIBROMIN (NEUROFIBROMATOSIS-RELATED PROTEIN NF-1).
 GN NFL.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; PRIMATES.
 RN [1]
 RP SEQUENCE OF 1-1370 AND 1392-2839 FROM N.A.
 RX MEDLINE; 92147138.
 RA MARCHUK D.A., SAULINO A., TAVAKKOL R., SWAROOP M., WALLACE M.R.,
 RA ANDERSEN L.B., MITCHELL A.L., GUTMANN D.H., BOGUSKI M., COLLINS F.S.;
 RL GENOMICS 11:931-940(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93090270.
 RA BERNARDS A., HAASE V.H., MURTHY A.E., MENON A., HANNIGAN G.E.,
 RA GUSELLA J.F.;
 RL DNA CELL BIOL. 11:727-734(1992).
 RN [3]
 RP SEQUENCE OF 335-1370 AND 1392-2839 FROM N.A.
 RX MEDLINE; 90335969.
 RA XU G., O'CONNELL P., VISKOCHIL D., CANTHON R., ROBERTSON M.,
 RA CULVER M., DUNN D., STEVENS J., GESTELAND R., WHITE R., WEISS R.;
 RL CELL 62:599-608(1990).
 RN [4]
 RP SEQUENCE OF 1096-1370 AND 1372-1590 FROM N.A.
 RX MEDLINE; 91029515.
 RA MARTIN G.A., VISKOCHIL D., BOLLAG G., MCCABE P.C., CROSIER W.J.,
 RA HAUBRUCK H., CONROY L., CLARK R., O'CONNELL P., CANTHON R.M.,
 RA INNIS M., MCCORMICK F.;
 RL CELL 63:843-849(1990).
 RN [5]
 RP SEQUENCE OF 1606-2709 FROM N.A., AND VARIANT PRO-1953.
 RX MEDLINE; 90304909.
 RA CANTHON R.M., WEISS R., XU G., VISKOCHIL D., CULVER M., STEVENS J.,
 RA ROBERTSON M., DUNN D., GESTELAND R., O'CONNELL P., WHITE R.;
 RL CELL 62:193-201(1990).
 RN [6]
 RP SEQUENCE OF 2230-2839 FROM N.A.
 RX MEDLINE; 90319792
 RA WALLACE M.R., MARCHUK D.A., ANDERSEN L.B., LETCHER R., ODEH H.M.,
 RA SAULINO A.M., FOUNTAIN J.W., BRERETON A., NICHOLSON J., MITCHELL A.L.,
 RA BROWNSTEIN B.H., COLLINS F.S.;
 RL SCIENCE 249:181-186(1990).
 RN [7]
 RP ERRATUM.
 RX MEDLINE; 91102559.
 RA WALLACE M.R., MARCHUK D.A., ANDERSEN L.B., COLLINS F.S.;
 RL SCIENCE 250:1749-1749(1990).
 RN [8]
 RP SEQUENCE OF 1168-1566 FROM N.A.
 RX MEDLINE; 92019823.
 RA NISHI T., LEE P.S., OKA K., LEVIN V.A., TANASE S., MORINO Y.,
 RA SAYA H.;
 RL ONCOGENE 6:1555-1559(1991).
 RN [9]
 RP SEQUENCE OF 1371-1391 FROM N.A.
 RX MEDLINE; 93109335.
 RA ANDERSEN L.B., BALLESTER R., MARCHUK D.A., CHANG E., GUTMANN D.H.,
 RA SAULINO A.M., CAMONIS J., WIGLER M., COLLINS F.S.;
 RL MOL. CELL. BIOL. 13:487-495(1993).

[10]
RN FUNCTION.
RX MEDLINE; 91029516.
RA BALLESTER R., MARCHUK D., BOGUSKI M.S., SAULINO A., LETCHER R.,
RA WIGLER M., COLLINS F.S.;
RL CELL 63:851-859(1990).
[11]
RN REVIEW ON VARIANTS.
RX MEDLINE; 95072625.
RA UPADHYAYA M., SHAW D.J., HARPER P.S.;
RL HUM. MUTAT. 4:83-101(1994).
[12]
RN REVIEW ON VARIANTS.
RX MEDLINE; 96422425.
RA HONG SHEN M., HARPER P.S., UPADHYAYA M.;
RL J. MED. GENET. 33:2-17(1996).
[13]
RN VARIANT GLU-1444.
RX MEDLINE; 92233464.
RA LI Y., BOLLAG G., CLARK R., STEVENS J., CONROY L., FULTS D., WARD K.,
RA FRIEDMAN E., SAMOWITZ W., ROBERTSON M., BRADLEY P., MCCORMICK F.,
RA WHITE R., CAWTHON R.;
RL CELL 69:275-281(1992).
[14]
RN VARIANTS MET-2164 AND ASN-2192.
RX MEDLINE; 93258316.
RA UPADHYAYA M., SHEN M., CHERRYSON A., FARNHAM J., MAYNARD J.,
RA HUSON S.M., HARPER P.S.;
RL HUM. MOL. GENET. 1:735-740(1992).
[15]
RN VARIANT HIS-1721--LEU-1733 DUPLICATION.
RX MEDLINE; 93304433.
RA TASSABEHJI M., STRACHAN T., SHARLAND M., COLLEY A., DONNAI D.,
RA HARRIS R., THAKER N.;
RL AM. J. HUM. GENET. 53:90-95(1993).
[16]
RN VARIANT MET-991 DEL.
RX MEDLINE; 94108439.
RA SHEN M.H., HARPER P.S., UPADHYAYA M.;
RL HUM. MOL. GENET. 2:1861-1864(1993).
[17]
RN VARIANT NF1 ASN-2387--PHE-2388 DEL.
RX MEDLINE; 94362704.
RA ABERNATHY C.R., COLMAN S.D., KOUSSEFF B.G., WALLACE M.R.;
RL HUM. MUTAT. 3:347-352(1994).
[18]
RN VARIANT NF1 ALA-2631.
RX MEDLINE; 96091873.
RA UPADHYAYA M., MAYNARD J., OSBORN M., HUSON S.M., PONDER M.,
RA PONDER B.A.J., HARPER P.S.;
RL J. MED. GENET. 32:706-710(1995).
[19]
RN VARIANT NF1 ARG-629.
RX MEDLINE; 96431167.
RA GASPARI P., D'AGRUMA L., DE CILLIS G.P., BALESTRAZZI P.,
RA MINGARELLI R., ZELANTE L.;
RL HUM. GENET. 97:492-495(1996).
[20]
RN VARIANT LS ARG-1035.
RX MEDLINE; 96400960.
RA WU R., LEGIUS E., ROBBERECHT W., DUMOULIN M., CASSIMAN J.-J.,
RA FRYNs J.-P.;
RL HUM. MUTAT. 8:51-56(1996).
[21]
RN VARIANTS NF1 ARG-844 AND PRO-898.
RX MEDLINE; 97295087.
RA MAYNARD J., KRAWCZAK M., UPADHYAYA M.;
RL HUM. GENET. 99:674-676(1997).
[22]
RN VARIANT NF1 ARG-1952.
RX MEDLINE; 97255969.
RA HUDSON J., WU C.L., TASSABEHJI M., SUMMERS E.M., SIMON S., SUPER M.,
RA DONNAI D., THAKER N.;

HUM. MUTAT. 9:366-367(1997).
[23]
RN VARIANT NF1 TRP-1611.
RX MEDLINE; 97442280.
RA UPADHYAYA M., MAYNARD J., OSBORN M., HARPER P.S.;
RL HUM. MUTAT. 10:248-250(1997).
-|- FUNCTION: STIMULATES THE GTPASE ACTIVITY OF RAS. NF1 SHOWS GREATER AFFINITY FOR RAS GAP, BUT LOWER SPECIFIC ACTIVITY. THUS IT MAY BE A REGULATOR OF RAS ACTIVITY.
-|- DISEASE: THIS PROTEIN IS ASSOCIATED WITH TYPE 1 NEUROFIBROMATOSIS (NF1) (ALSO CALLED VON RECKLINGHAUSEN SYNDROME), THE MOST FREQUENT INHERITED GENETIC DISEASE (ABOUT 1 IN 3000). IT EXHIBITS FULL PENETRANCE AND HIGH MUTATION RATE WITH 30 TO 50% OF NF1 PATIENTS REPRESENTING A NEW MUTATION. AMONG THE MANY CLINICAL FEATURES OF NF1 ARE PATCHES OF SKIN PIGMENTATION (CAFE-AU-LAIT SPOTS), LISCH NODULES OF THE IRIS PERIPHERAL, PERIPHERAL NERVOUS SYSTEM ASSOCIATED TUMORS AND FIBROMATOUS SKIN TUMORS. THE DISEASE DEMONSTRATES A HIGH DEGREE OF PENETRANCE BY AGE 5 YEARS.
-|- DISEASE: DEFECTS IN NF1 ARE ASSOCIATED WITH WATSON SYNDROME (WS). WS IS CHARACTERIZED BY THE PRESENCE OF PULMONARY STENOSIS, CAFE-AU-LAIT SPOTS, AND MENTAL RETARDATION. WS IS CONSIDERED AS AN ATYPICAL FORM OF NF1.
-|- DISEASE: DEFECTS IN NF1 ARE ASSOCIATED WITH LEOPARD, AN AUTOSOMAL DOMINANT SYNDROME. LEOPARD (LS) IS AN ACRONYM FOR THE FEATURES OF THIS SYNDROME: L-LENTIGINES ("FRECKLES"), E-ELECTROCARDIOGRAPHIC ABNORMALITIES, O-OCULAR HYPERTELOISM, P-PULMONARY STENOSIS, A-ABNORMALITIES OF GENITALIA, R-RETARDATION OF GROWTH, AND D-DEAFNESS (SENSORINEURAL). THE MAIN FEATURES OF THE SYNDROME ARE MULTIPLE LENTIGINES IN COMBINATION WITH A CONGENITAL HEART MALFORMATION (PULMONARY STENOSIS, SUBVALVULAR MUSCULAR AORTIC STENOSIS). A CLINICAL OVERLAP EXISTS BETWEEN LS, NF1 AND WS.
-|- ALTERNATIVE PRODUCTS: TWO FORMS OF NEUROFIBROMIN ARE PRODUCED BY ALTERNATIVE SPLICING OF THE NF1 GENE; TYPE I AND TYPE II (SHOWN HERE) ONLY DIFFERS IN THE INSERTION OF A 21 RESIDUES SEGMENT.
-|- SIMILARITY: TO OTHER RAS GTPASE-ACTIVATING PROTEINS.
EMBL; M82814; G189165; -
EMBL; M89914; G292354; -
EMBL; M38116; G494225; -
EMBL; M38108; G494225; JOINED.
EMBL; M38109; G494225; JOINED.
EMBL; M38110; G494225; JOINED.
EMBL; M38111; G494225; JOINED.
EMBL; M38112; G494225; JOINED.
EMBL; M38113; G494225; JOINED.
EMBL; M38114; G494225; JOINED.
EMBL; M38106; G189170; -
EMBL; M61213; G189163; -
EMBL; M38107; G189172; ALT_SEQ.
EMBL; M60496; G189158; -
EMBL; M60915; G189161; -
EMBL; S51751; G362288; -
EMBL; D12625; G219940; ALT_SEQ.
PIR; A35222; A35222.
PIR; A35605; A35605.
PIR; A35879; A35879.
PIR; JC1277; JC1277.
DR MIN; 162200; -
DR MIN; 193520; -
DR MIN; 151100; -
DR PROSITE; PS00509; RAS_GTPASE_ACTIV_1; 1.
DR PROSITE; PS00018; RAS_GTPASE_ACTIV_2; 1.
KW GTPASE ACTIVATION; ALTERNATIVE SPLICING; ANTI-ONCOGENE;
...
Note: remainder of annotations omitted.
Query Match 70.3%; Score 45; DB 1; Length 2839;
Best Local Similarity 60.0%; Pred. No. 9.01e-00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 1547 GPPHKKPVAD 1556
:||||:||||
Qy 1 APPENDPVED 10

```

RESULT 8
ID NF1_MOUSE STANDARD; PRT; 2841 AA.
AC Q04690; Q61956; Q61957;
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DE NEUROFIBROMIN (NEUROFIBROMATOSIS-RELATED PROTEIN NF-1).
GN NF1
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-BRAIN;
RX MEDLINE; 93357730.
RA BERNARDS A., SNIDJERS A.J., HANNIGAN G.E., MURTHY A.E., GUSELLA J.F.;
RL HUM. MOL. GENET. 2:645-650(1993).
[2]
RP SEQUENCE OF 1178-1555 FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE; 95047432.
RA MANTANI A., MAKASUGI S., YOKOTA Y., ABE K., USHIO Y., YAMAMURA K.;
RL GENE 148:245-251(1994).
[3]
RP SEQUENCE OF 1950-2568 FROM N.A.
RX MEDLINE; 90384569.
RA BUCHBERG A.M., CLEVELAND L.S., JENKINS N.A., COPELAND N.G.;
RL NATURE 347:291-294(1990).
CC -!- FUNCTION: STIMULATES THE GTPASE ACTIVITY OF RAS. NF1 SHOWS GREATER
AFFINITY FOR RAS GAP, BUT LOWER SPECIFIC ACTIVITY. THUS IT MAY BE
A REGULATOR OF RAS ACTIVITY.
CC -!- TISSUE SPECIFICITY: TYPE I IS EXPRESSED PREDOMINANTLY IN BRAIN,
SPINAL CORD AND TESTIS. TYPE II IS EXPRESSED PREDOMINANTLY IN
ADRENAL GLAND, KIDNEY, OVARY AND LUNG. TYPE III IS EXPRESSED
PREDOMINANTLY IN ADRENAL GLAND AND TYPE IV IS EXPRESSED
MAINLY IN THE TESTIS.
CC -!- ALTERNATIVE PRODUCTS: FOUR FORMS OF THE PROTEIN (TYPES I, II,
III AND IV) ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME
GENE. THE SEQUENCE SHOW HERE IS THAT OF TYPE II.
CC -!- SIMILARITY: TO OTHER RAS GTPASE-ACTIVATING PROTEINS.
DR EMBL; L10369; G309451; -
DR EMBL; L10367; G309451; JOINED.
DR EMBL; L10368; G309451; JOINED.
DR EMBL; L10370; G309453; -
DR EMBL; X54924; G930191; -
DR EMBL; D30730; G577638; -
DR EMBL; D30731; G577640; -
DR MGI; MGI:97306; NF1.
DR PROSITE; PS00509; RAS_GTPASE_ACTIV_1; 1.
DR PROSITE; PS50018; RAS_GTPASE_ACTIV_2; 1.
KW GTPASE ACTIVATION; ALTERNATIVE SPLICING.
FT DOMAIN 1237 1453
FT VARSPLIC 1373 1393
FT VARSPLIC 1394 1406
FT VARSPLIC 1407 2841
FT VARSPLIC 2841 2841 MISSING (IN TYPE III AND TYPE IV).
SQ SEQUENCE 2841 AA; 319591 MW; A7AA76F4 CRC32;

Query Match 70.3%; Score 45; DB 1; Length 2841;
Best Local Similarity 60.0%; Pred. No. 9.01e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 1549 GPEHKKPVAD 1558
:|||||:
QY 1 APPEDNPVED 10

RESULT 9
ID PAHO_LARAR STANDARD; PRT; 36 AA.
AC P41337;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DE PANCREATIC HORMONE PRECURSOR (PANCREATIC POLYPEPTIDE) (PP).
OS GALLUS GALLUS (CHICKEN), AND MELEAGRIS GALLOPAVO (COMMON TURKEY).

```

```

DE PANCREATIC HORMONE (PANCREATIC POLYPEPTIDE) (PP).
OS LARUS ARGENTATUS (HERRING GULL).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC CHARADRIIFORMES.
[1]
RP SEQUENCE.
RC TISSUE=PANCREAS;
RX MEDLINE; 94229519.
RA BARTON C.L., SHAW C., HALTON D.W., THIM L.;
RL GEN. COMP. ENDOCRINOL. 93:255-259(1994).
CC -!- FUNCTION: PANCREATIC HORMONE IS SYNTHESIZED IN PANCREATIC ISLETS
OF LANGERHANS AND ACTS AS A REGULATOR OF PANCREATIC AND
GASTROINTESTINAL FUNCTIONS.
CC -!- SIMILARITY: BELONGS TO THE PP FAMILY (PANCREATIC HORMONE,
NEUROPEPTIDE Y, PEPTIDE YY).
DR PROSITE; PS00265; PANCREATIC_HORMONE; 1.
KW HORMONE; AMIDATION; PANCREAS.
FT MOD_RES 36
FT SEQUENCE 36 AA; 4237 MW; 2AF9A777 CRC32;

Query Match 68.8%; Score 44; DB 1; Length 36;
Best Local Similarity 66.7%; Pred. No. 1.47e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 8 PGDDAPVED 16
|:|:|
QY 2 PPEDNPVED 10

RESULT 10
ID PAHO_SYRCA STANDARD; PRT; 36 AA.
AC P11967;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT 01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE PANCREATIC HORMONE (PANCREATIC POLYPEPTIDE) (PP).
OS STRUTHIO CAMELUS (OSTRICH).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES;
OC PALAEOGNATHAE; STRUTHIONIFORMES.
[1]
RP SEQUENCE.
RX MEDLINE; 87307111.
RA LITTHAUER D., OELOFSEN W.;
RL INT. J. PEPT. PROTEIN RES. 29:739-745(1987).
CC -!- FUNCTION: PANCREATIC HORMONE IS SYNTHESIZED IN PANCREATIC ISLETS
OF LANGERHANS AND ACTS AS A REGULATOR OF PANCREATIC AND
GASTROINTESTINAL FUNCTIONS.
CC -!- SIMILARITY: BELONGS TO THE PP FAMILY (PANCREATIC HORMONE,
NEUROPEPTIDE Y, PEPTIDE YY).
DR PIR; A28578; A28578.
DR HSSP; P01306; 1PPT.
DR PROSITE; PS00265; PANCREATIC_HORMONE; 1.
KW HORMONE; AMIDATION; PANCREAS.
FT MOD_RES 36
FT SEQUENCE 36 AA; 4209 MW; 45F8AEDB CRC32;

Query Match 68.8%; Score 44; DB 1; Length 36;
Best Local Similarity 66.7%; Pred. No. 1.47e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 8 PGDDAPVED 16
|:|:|
QY 2 PPEDNPVED 10

RESULT 11
ID PAHO_CHICK STANDARD; PRT; 80 AA.
AC P01306;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE PANCREATIC HORMONE PRECURSOR (PANCREATIC POLYPEPTIDE) (PP).
OS GALLUS GALLUS (CHICKEN), AND MELEAGRIS GALLOPAVO (COMMON TURKEY).

```

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
 OC GALLIFORMES.
 RN [1]
 RC SPECIES=CHICKEN: TISSUE=LIVER;
 RX MEDLINE; 93366173.
 RA NATA K., SUGIMOTO T., KOHRI K., HIDAKA H., HATTORI E., YAMAMOTO H.,
 RA YONEKURA H., OKAMOTO H.;
 RL GENE 130:183-189(1993).
 RN [2]
 RP SEQUENCE OF 26-61.
 RC SPECIES=CHICKEN;
 RX MEDLINE; 76069270.
 RA KIMMEL J.R., HAYDEN L.J., POLLOCK H.G.;
 RL J. BIOL. CHEM. 250:9369-9376(1975).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (1.4 ANGSTROMS).
 RC SPECIES=TURKEY;
 RX MEDLINE; 84179397.
 RA GLOVER I., MANEUF I., PITTS J., WOODS S., MOSS D., TICKLE I.,
 RA BLUNDELL T.L.;
 RL BIOPOLYMERS 22:293-304(1983).
 CC -!- FUNCTION: PANCREATIC HORMONE IS SYNTHESIZED IN PANCREATIC ISLETS
 OF LANGERHANS AND ACTS AS A REGULATOR OF PANCREATIC AND
 GASTROINTESTINAL FUNCTIONS.
 CC -!- SIMILARITY: BELONGS TO THE PP FAMILY (PANCREATIC HORMONE,
 NEUROPEPTIDE Y, PEPTIDE YY).
 DR EMBL; D13761; G391634; -.
 DR EMBL; D13760; G391646; -.
 DR PIR; A01575; PCCH.
 DR PDB; 1PPT; 15-OCT-91.
 DR PROSITE; PS00265; PANCREATIC HORMONE; 1.
 KW HORMONE; CLEAVAGE ON PAIR OF BASIC RESIDUES; PANCREAS; SIGNAL;
 KW AMIDATION; 3D-STRUCTURE.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 61 PANCREATIC HORMONE.
 FT MOD_RES 61 61 AMIDATION (G-62 PROVIDE AMIDE GROUP).
 FT CONFLICT 47 48 ND -> DN (IN REF. 2).
 FT TURN 35 36
 FT HELIX 39 56
 FT TURN 57 58
 SQ SEQUENCE 80 AA; 8773 MW; DE642B29 CRC32;
 Query Match 68.8%; Score 44; DB 1; Length 80;
 Best Local Similarity 66.7%; Pred. No. 1.47e+01;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Db 33 PGDDAPVED 41
 QY 2 PPDNPVED 10
 RESULT 12
 ID YLC3_YEREN STANDARD; PRT; 122 AA.
 AC P21308;
 DT 01-MAY-1991 (REL. 18, CREATED)
 DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
 DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 13.7 KD PROTEIN IN LCRC 3'REGION (ORF3).
 OS YERSINIA ENTEROCOLITICA.
 OG PLASMID PYV.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 OC ENTEROBACTERIACEAE.
 RN [1]
 RC SEQUENCE FROM N.A.
 RX STRAIN=SEROTYPED O:3;
 RX MEDLINE; 90264308.
 RA VIITANEN A.-M., TOIVANEN P., SKURNIK M.;
 RL J. BACTERIOL. 172:3152-3162(1990).
 CC -!- INDUCTION: TEMPERATURE SEEMS TO PLAY THE MAJOR ROLE IN REGULATION
 OF TRANSCRIPTION OF THE LCRC-CONTAINING OPERON OF PYV, WHEREAS
 CA(2+) CONCENTRATION HAS ONLY A MODERATE EFFECT AT 37 DEGREES
 CELSIUS, AND NO EFFECT AT ROOM TEMPERATURE.

DR EMBL; M32097; -; NOT_ANNOTATED_CDS.
 DR PIR; D35392; D35392.
 KW HYPOTHETICAL PROTEIN; PLASMID.
 SQ SEQUENCE 122 AA; 13741 MW; 0E8B386D CRC32;
 Query Match 68.8%; Score 44; DB 1; Length 122;
 Best Local Similarity 62.5%; Pred. No. 1.47e+01;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 34 PPDGHPVE 41
 QY 2 PPDNPVE 9
 RESULT 13
 ID NEUM_CARAU STANDARD; PRT; 213 AA.
 AC P17691;
 DT 01-AUG-1990 (REL. 15, CREATED)
 DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE NEUROMODULIN (AXONAL MEMBRANE PROTEIN GAP-43) (PP46) (B-50) (PROTEIN
 FL) (CALMODULIN-BINDING PROTEIN P-57).
 OS CARASSIUS AURATUS (GOLDFISH).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES; GNATHOSTOMATA;
 OC OSTEICHTHYES; ACTINOPTERYGII; CYPRINIFORMES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 90380372.
 RA LABATE M.E., SKENE J.H.P.;
 RL NEURON 3:299-310(1989).
 CC -!- FUNCTION: THIS PROTEIN IS ASSOCIATED WITH NERVE GROWTH. IT IS A
 MAJOR COMPONENT OF THE MOTILE "GROWTH CONES" THAT FORM THE TIPS
 OF ELONGATING AXONS.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF GROWTH CONE AND
 SYNAPTIC PLASMA MEMBRANES.
 CC -!- PTM: PHOSPHORYLATION OF THIS PROTEIN BY A PROTEIN KINASE C IS
 SPECIFICALLY CORRELATED WITH CERTAIN FORMS OF SYNAPTIC PLASTICITY.
 CC -!- BINDS CALMODULIN WITH A GREATER AFFINITY IN THE ABSENCE OF CA++
 THAN IN ITS PRESENCE.
 DR EMBL; M26250; G212954; -.
 DR PIR; JQ0075; JQ0075.
 DR PROSITE; PS00412; NEUROMODULIN_1; 1.
 DR PROSITE; PS00413; NEUROMODULIN_2; 1.
 KW NEURONE; PHOSPHORYLATION; MEMBRANE; GROWTH REGULATION;
 KW CALMODULIN-BINDING; LIPOPROTEIN.
 FT DOMAIN 1 4 IMPORTANT FOR MEMBRANE BINDING.
 FT LIPID 3 3 PALMITATE (PROBABLE).
 FT LIPID 4 4 PALMITATE (PROBABLE).
 FT MOD_RES 42 42 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 FT DOMAIN 44 52 CALMODULIN-BINDING (PROBABLE).
 FT DOMAIN 155 170 POLY-GLU.
 SQ SEQUENCE 213 AA; 23569 MW; BFAC407B CRC32;
 Query Match 68.8%; Score 44; DB 1; Length 213;
 Best Local Similarity 55.6%; Pred. No. 1.47e+01;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 Db 79 SPSEKPEV 87
 QY 1 APPEDNPVE 9
 RESULT 14
 ID TRT3_CHICK STANDARD; PRT; 262 AA.
 AC P12620; P12621; P12619; P12618;
 DT 01-OCT-1989 (REL. 12, CREATED)
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE TROPONIN T, FAST SKELETAL MUSCLE ISOFORMS.
 OS GALLUS GALLUS (CHICKEN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
 OC GALLIFORMES.
 RN [1]

W P S R E H
***** (TW)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu May 13 15:28:10 1999; MasPar time 2.96 Seconds
Tabular output not generated. 126.733 Million cell updates/sec

Title: >US-09-040-485-6
Description: (1-10) from US09040485.pap
Perfect Score: 64
Sequence: 1 APPEDNPVED 10

Scoring table: PAM 150
Gap 15

Searched: 116738 seqs, 37463448 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r58
1.p1r1 2.p1r2 3.p1r3 4.p1r4

Statistics: Mean 21.139; Variance 27.501; scale 0.769

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	64	100.0	757	2	I38423	aspartyl beta-hydroxy
2	50	78.1	704	1	S60117	protein kinase C (EC
3	47	73.4	129	2	A70970	hypothetical protein
4	46	71.9	562	2	JC5182	serine C-palmitoyl
5	45	70.3	86	2	S76775	hypothetical protein
6	45	70.3	97	2	S11755	hypothetical protein
7	45	70.3	258	2	C70885	probable dehydrogenas
8	45	70.3	339	2	B36868	copB homolog - Xantho
9	45	70.3	399	2	I78852	neurofibromatosis pro
10	45	70.3	928	1	R8HU	retinoblastoma-associ
11	45	70.3	932	2	B9552	leucyl-tRNA synthetas
12	45	70.3	2818	2	B5282	neurofibromatosis-rel
13	45	70.3	2820	2	JC5196	neurofibromin I - rat
14	45	70.3	2825	2	I54352	neurofibromin - mouse
15	44	68.8	36	2	A28578	pancreatic hormone -
16	44	68.8	80	1	PCCH	pancreatic hormone pr
17	44	68.8	122	2	D35392	hypothetical protein
18	44	68.8	213	2	JQ0075	neuromodulin - goldfi
19	44	68.8	257	2	B31957	tropoinin T, skeletal
20	44	68.8	259	2	F69311	conserved hypothetical
21	44	68.8	263	2	C31957	tropoinin T, skeletal
22	44	68.8	304	2	A48174	beta-lactamase (EC 3.
23	44	68.8	376	2	S12730	actin - California se

24	44	68.8	395	2	S36029	hypothetical protein
25	44	68.8	979	2	A35913	regulatory factor X -
26	43	67.2	176	2	S06638	calpastatin - bovine
27	43	67.2	263	2	C63044	hypothetical protein
28	43	67.2	307	2	A36885	bo-type ubiquinol oxi
29	43	67.2	491	2	A49993	glycylpeptide N-tetra
30	43	67.2	506	1	FOLJG5	gag polyprotein - sim
31	43	67.2	507	2	S04237	gag polyprotein - sim
32	43	67.2	966	2	S25365	CYC8 protein - yeast
33	43	67.2	976	2	S40697	processing endoprotei
34	43	67.2	1257	2	A41060	neural cell adhesion
35	43	67.2	1259	2	S36126	neural cell adhesion
36	43	67.2	1260	2	S05479	neural cell adhesion
37	42	65.6	230	2	S72714	Lepb170.F2.64 protei
38	42	65.6	313	2	S26838	modulation protein no
39	42	65.6	670	2	G71251	probable DNA helicase
40	42	65.6	837	2	D71027	hypothetical protein
41	42	65.6	1133	2	S54496	probable membrane pro
42	42	65.6	1148	2	A71446	hypothetical protein
43	42	65.6	1836	2	JS0648	sodium channel alpha
44	42	65.6	1836	2	IS1964	sodium channel alpha
45	42	65.6	2638	1	A42545	genome polyprotein -

ALIGNMENTS

RESULT 1
ENTRY I38423 #type complete
TITLE aspartyl beta-hydroxylase - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 29-May-1998 #sequence_revision 29-May-1998 #text_change 10-Jul-1998
ACCESSIONS I38423
REFERENCE I38423
#authors Koriath, F.; Gieffers, C.; Frey, J.
#journal Gene (1994) 150:395-399
#title Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase.
#cross-references MUID:95121937
#accession I38423
##status preliminary; translated from GB/EMBL/DBDJB
##molecule_type mRNA
##residues 1-757 ##label RES
##cross-references EMBL:U03109; NID:9458031; PID:9458032
CLASSIFICATION #superfamily peptide-aspartate beta-dioxygenase;
tetrairicopeptide repeat homology
FEATURE
54-75 #domain transmembrane #status predicted #label TRM
SUMMARY #length 757 #molecular-weight 85496 #checksum 2143

Query Match 100.0%; Score 64; DB 2; Length 757;
Best Local Similarity 100.0%; Pred. No. 1.75e-03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 281 APPEDNPVED 290
Qy 1 APPEDNPVED 10

RESULT 2
ENTRY S60117 #type complete
TITLE protein kinase C (EC 2.7.1.1) TPA-1A - Caenorhabditis elegans
CONTAINS protein kinase C TPA-1B
ORGANISM #formal_name Caenorhabditis elegans
DATE 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 05-Sep-1997
ACCESSIONS S60117
REFERENCE S60117
#authors Sano, T.; Tabuse, Y.; Nishiwaki, K.; Miwa, J.
#journal J. Mol. Biol. (1995) 251:477-485
#title The tpa-1 gene of Caenorhabditis elegans encodes two proteins similar to Ca(2+)-independent protein kinase Cs: evidence by complete genomic and complementary DNA sequences of the

```

tpa-1 gene.
#accession S60117
#molecule_type DNA
##residues 1-704 ##label SAN
##cross-references EMBL:D49525; NID:g1217583; PID:d1009081; PID:g1217584
GENETICS
#gene tpa-1
#map_position IV
#introns 9/2; 45/1; 93/3; 128/1; 153/1; 226/3; 270/1; 446/3; 498/2; 655/3
FUNCTION
#description catalyzes the formation of peptidyl-serine-phosphate or
peptidyl-threonine-phosphate using ATP
#note activity is calcium-independent, phospholipid-dependent, and
activated by diacylglycerol and by tumor-promoting phorbol
esters
CLASSIFICATION #superfamily protein kinase C delta; protein kinase C
zinc-binding repeat homology; protein kinase homology
alternative splicing; ATP; duplication; phorbol ester
binding; phospholipid binding; phosphotransferase;
serine/threonine-specific protein kinase; zinc
FEATURE
138-704 #product protein kinase C TPA-1B #status predicted
#label MAT2\
#region pseudophosphorylation motif\
#domain protein kinase C zinc-binding repeat homology
#label K21\
#domain protein kinase C zinc-binding repeat homology
#label K22\
#domain protein kinase homology #label KIN\
#region protein kinase ATP-binding motif\
#binding_site zinc (His, Cys, Cys, Cys) #status
predicted\
179,182,204,207 #binding_site zinc (Cys, Cys, His, Cys) #status
predicted\
238,268,271,287 #binding_site zinc (His, Cys, Cys, Cys) #status
predicted\
251,254,276,279 #binding_site zinc (Cys, Cys, His, Cys) #status
predicted\
404,423,499,501 #active_site Lys, Glu, Asp, Lys #status predicted
SUMMARY #length 704 #molecular-weight 80298 #checksum 7955
Query Match 78.1%; Score 50; DB 1; Length 704;
Best Local Similarity 70.0%; Pred. No. 2.17e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 127 APDDHPVED 136
|| :|||
QY 1 APPEDNPVED 10
RESULT 3
ENTRY #type complete
TITLE hypothetical protein RV3354 - Mycobacterium tuberculosis
(strain H37RV)
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS A70970
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry
III, C.E.; Tekata, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; 393:537-544
Nature (1998)
#journal Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#title

```

```

#cross-references MUID:98295987
#accession A70970
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
##residues 1-129 ##label COL
##cross-references GB:AL009198; GB:AL123456; NID:g3242262; PID:e1202271;
PID:g2661632
#experimental_source strain H37RV
GENETICS
#gene RV3354
SUMMARY #length 129 #molecular-weight 12988 #checksum 6975
Query Match 73.4%; Score 47; DB 2; Length 129;
Best Local Similarity 60.0%; Pred. No. 8.79e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 28 APAQANPVD 37
||: |||
QY 1 APPEDNPVED 10
RESULT 4
ENTRY #type complete
TITLE serine C-palmitoyltransferase (EC 2.3.1.50) Lcb2 chain -
Yeast (Kluyveromyces fragilis)
ALTERNATE_NAMES 3-ketosphinganine synthetase; SPT
ORGANISM #formal_name Kluyveromyces fragilis
DATE 16-Feb-1997 #sequence_revision 27-Feb-1997 #text_change
13-Sep-1998
ACCESSIONS JC5182
REFERENCE JC5180
#authors Nagiec, M.M.; Lester, R.L.; Dickson, R.C.
#journal Gene (1996) 177:237-241
#title Sphingolipid synthesis: Identification and characterization
of mammalian cDNAs encoding the Lcb2 subunit of serine
palmitoyltransferase.
#accession JC5182
#status preliminary; nucleic acid sequence not shown
#molecule_type mRNA
##residues 1-562 ##label NAG
##cross-references GB:U15646; NID:g1001948; PID:g1001949
COMMENT This enzyme catalyzes the irreversible reaction of condensation of
serine and palmitoyl-CoA to yield 3-ketosphinganine.
CLASSIFICATION #superfamily serine C-palmitoyltransferase chain LCB2
acyltransferase
KEYWORDS 319-374
FEATURE #region Lcb2 protein signature
SUMMARY #length 562 #molecular-weight 63004 #checksum 5581
Query Match 71.9%; Score 46; DB 2; Length 562;
Best Local Similarity 77.8%; Pred. No. 1.38e+01;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 13 PPEDIPLED 21
||| |||
QY 2 PPEDNPVED 10
RESULT 5
ENTRY #type complete
TITLE hypothetical protein - Synecocystis sp. (strain PCC 6803)
ORGANISM #formal_name Synecocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998
ACCESSIONS S76775
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;

```

```

#journal      DNA Res. (1996) 3:109-136
#title       Sequence analysis of the genome of the unicellular
#            cyanobacterium Synechocystis sp. PCC6803. II. Sequence
#            determination of the entire genome and assignment of
#            potential protein-coding regions.
#cross-references MUID:97061201
#accession     S76775
##status      preliminary
##molecule_type DNA
##residues     1-86 ##label KAN
##cross-references EMBL:D90916; GB:AB001339; NID:g1653715; PID:d1019420;
##note        the nucleotide sequence was submitted to the EMBL Data
##            Library, June 1996
SUMMARY       #length 86 #molecular-weight 9909 #checksum 7709

Query Match   70.3%; Score 45; DB 2; Length 86;
Best Local Similarity 62.5%; Pred. No. 2.17e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 49 PDDTPVEE 56
   ||| |||
QY 3 PEDNPVED 10

RESULT 6
ENTRY 6
TITLE hypothetical protein - Plasmodium vivax (fragment)
ORGANISM #formal_name Plasmodium vivax
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
09-Sep-1997
ACCESSIONS S11755
REFERENCE S11755
#authors Campbell, J.R.; Franke, E.D.
#submission submitted to the EMBL Data Library, April 1989
#description A repetitive element found in Plasmodium vivax DNA.
#accession S11755
##molecule_type DNA
##residues 1-97 ##label CAM
##cross-references EMBL:X15129; NID:g10086; PID:g10087
SUMMARY #length 97 #checksum 7475

Query Match   70.3%; Score 45; DB 2; Length 97;
Best Local Similarity 71.4%; Pred. No. 2.17e+01;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 13 PPEENPI 19
   ||| |||
QY 2 PEDNPV 8

RESULT 7
ENTRY 7
TITLE probable dehydrogenase - Mycobacterium tuberculosis (strain
#formal_name Mycobacterium tuberculosis
#17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS C70885
REFERENCE A70500
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
III, C.E.; Tekle, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsbey, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrall, B.G.
Nature (1998) 393:537-544
Deciphering the biology of Mycobacterium tuberculosis from

#journal
#title
the complete genome sequence.
#cross-references MUID:98295987
#accession     C70885
##status      preliminary; nucleic acid sequence not shown;
##molecule_type DNA
##residues     1-258 ##label COL
##cross-references GB:AL008883; GB:AL123456; NID:g3261490; PID:ell172949;
##experimental_source strain H37Rv
GENETICS      RV2857C
#gene         #length 258 #molecular-weight 26804 #checksum 1750
SUMMARY

Query Match   70.3%; Score 45; DB 2; Length 258;
Best Local Similarity 60.0%; Pred. No. 2.17e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 93 SPEDNLIEN 102
   ||||| |||
QY 1 APPEDNPVED 10

RESULT 8
ENTRY 8
TITLE copB homolog - Xanthomonas campestris
ALTERNATE_NAMES hypotheticalprotein 2
ORGANISM #formal_name Xanthomonas campestris
DATE 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change
09-Sep-1997
ACCESSIONS B36868
REFERENCE B36868
#authors Lee, Y.A.; Henderson, M.; Panopoulos, N.J.; Schroth, M.N.
#journal J. Bacteriol. (1994) 176:173-188
#title Molecular cloning, chromosomal mapping, and sequence analysis
of copper resistance genes from Xanthomonas campestris pv.
juglandis: homology with small blue copper proteins and
multicopper oxidase.
#accession B36868
##status      preliminary
##molecule_type DNA
##residues     1-339 ##label LEE
##cross-references GB:L19222; NID:g349160; PID:g461142
SUMMARY #length 339 #molecular-weight 37041 #checksum 1404

Query Match   70.3%; Score 45; DB 2; Length 339;
Best Local Similarity 60.0%; Pred. No. 2.17e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 121 PPSDHPVHD 130
   ||| ||| |
QY 1 APPEDNPVED 10

RESULT 9
ENTRY 9
TITLE neurofibromatosis protein type 1 - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
29-Aug-1997
ACCESSIONS I78852
REFERENCE I58356
#authors Nishi, T.; Lee, P.S.; Oka, K.; Levin, V.A.; Tanase, S.;
Morino, Y.; Saya, H.
#journal Oncogene (1991) 6:1555-1559
#title Differential expression of two types of the neurofibromatosis
type 1 (NF1) gene transcripts related to neuronal
differentiation.
#cross-references MUID:92019823
#accession I78852
##status      preliminary; translated from GB/EMBL/DDBJ
##molecule_type mRNA
##residues     1-399 ##label RES

```

##cross-references GB:M60915; NID:g189159; PID:g189161

GENETICS

#gene GDB:NFI
##cross-references GDB:120231; OMIM:162200
#map_position 17q11.2-17q11.2
CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology
FEATURE
68-303 #domain ras-specific GAP catalytic domain homology
#label GAP
SUMMARY #length 399 #checksum 4004

Query Match 70.3%; Score 45; DB 2; Length 399;
Best Local Similarity 60.0%; Pred. No. 2.17e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 380 GPEHKPVAD 389

QY 1 :||| :|| |
1 :ADPNPVED 10

RESULT 10

ENTRY RBHU #type complete
TITLE retinoblastoma-associated protein - human
ALTERNATE_NAMES retinoblastoma susceptibility protein
ORGANISM #formal_name Homo sapiens #common_name man
DATE 30-Jun-1987 #sequence_revision 30-Jun-1990 #text_change
05-Sep-1997
ACCESSIONS JS0276; A03152; A91613; A39947; A49887; I54364; I58362;
I78863; I78866; I78872; I78873; A35590
REFERENCE JS0276
#authors Lee, W.H.; Shew, J.Y.; Hong, F.D.; Sery, T.W.; Donoso, L.A.;
Young, L.J.; Bookstein, R.; Lee, E.Y.H.P.
#journal Nature (1987) 329:642-645
#title The retinoblastoma susceptibility gene encodes a nuclear
phosphoprotein associated with DNA binding activity.

##cross-references MUID:88014238
#accession JS0276
##molecule_type mRNA
##residues 1-928 #label LE1
##cross-references GB:M28419; NID:g190962; PID:g190963
#note this sequence has two possible initiation sites, 1-Met
and 113-Met

REFERENCE

A03152
#authors Lee, W.H.; Bookstein, R.; Hong, F.; Young, L.J.; Shew, J.Y.;
Lee, E.Y.H.P.
#journal Science (1987) 235:1394-1399
#title Human retinoblastoma susceptibility gene: cloning,
identification, and sequence.
#cross-references MUID:87149066
#accession A03152
##molecule_type mRNA
##residues 113-116, 'LLSYRNTY', 125-332, 'R', 334-367, 'I', 369-928
#label LE2

##cross-references GB:M15400; NID:g190958; PID:g190959

REFERENCE

A91613
#authors McGee, T.L.; Yandell, D.W.; Dryja, T.P.
#journal Gene (1989) 80:119-128
#title Structure and partial genomic sequence of the human
retinoblastoma susceptibility gene.
#cross-references MUID:90006771
#accession A91613

##molecule_type DNA
##residues 1-928 #label MCG

##cross-references GB:M27845; GB:L11910; NID:g292420; PID:g292421
#note the authors translated the codon GAA for residue 559 as
Gly

REFERENCE

A39947
#authors Friend, S.H.; Horowitz, J.M.; Gerber, M.R.; Wang, X.F.;
Bogenmann, E.; Li, F.P.; Weinberg, R.A.
#journal Proc. Natl. Acad. Sci. U.S.A. (1987) 84:9059-9063
#title Deletions of a DNA sequence in retinoblastomas and
mesenchymal tumors: organization of the sequence and its
encoded protein.

##cross-references MUID:88097427

#accession A39947

##molecule_type mRNA

##residues 1-928 #label FRI
##cross-references GB:M33647; GB:J02994; NID:g190945; PID:g190946
REFERENCE A4987

#authors

T'Ang, A.; Wu, K.J.; Hashimoto, T.; Liu, W.Y.; Takahashi, R.;
Shi, X.H.; Mihara, K.; Zhang, F.H.; Chen, Y.Y.; Du, C.;
Qian, J.; Lin, Y.G.; Murphree, A.L.; Qiu, W.R.; Thompson,
T.; Benedict, W.F.; Fung, Y.K.T.

#journal

Oncogene (1989) 4:401-407

#title

Genomic organization of the human retinoblastoma gene.

#cross-references

MUID:89239454

#accession

A44987

#molecule_type

DNA

#residues

1-46 #label TAA

#cross-references

EMBL:X16439; NID:g35894; PID:g35895

REFERENCE

I54364

#authors

Lohmann, D.R.; Brandt, B.; Hopping, W.; Passarge, E.;
Horsthemke, B.

#journal

Hum. Mol. Genet. (1994) 3:2187-2193

#title

Spectrum of small length germline mutations in the Rb1 gene.

#accession

I54364

#status

preliminary; translated from GB/EMBL/DBJ

#molecule_type

DNA

#residues

128-133 #label LOH

#cross-references

GB:L49209; NID:g1088286; PID:g1088287

REFERENCE

I58362

#authors

Hogg, A.; Onadim, Z.; Baird, P.N.; Cowell, J.K.

#journal

Oncogene (1992) 7:1445-1451

#title

Detection of heterozygous mutations in the Rb1 gene in
retinoblastoma patients using single-strand conformation
polymorphism analysis and polymerase chain reaction
sequencing.

#cross-references

MUID:92319557

#accession

I58362

#status

preliminary; translated from GB/EMBL/DBJ

#molecule_type

DNA

#residues

377-394 #label RE5

#cross-references

GB:L41900; NID:g793969; PID:g793972

#accession

I78866

#status

preliminary; translated from GB/EMBL/DBJ

#molecule_type

DNA

#residues

655-671 #label RE4

#cross-references

GB:L41907; NID:g801729; PID:g801730

#accession

I78872

#status

preliminary; translated from GB/EMBL/DBJ

#molecule_type

DNA

#residues

889-904 #label RES

#cross-references

GB:L41913; NID:g794008; PID:g794009

#accession

I78873

#status

preliminary; translated from GB/EMBL/DBJ

#molecule_type

DNA

#residues

906-928 #label RE3

#cross-references

GB:L41914; NID:g794010; PID:g794011

REFERENCE

A35590

#authors

Lee, E.Y.H.P.; Bookstein, R.; Young, L.J.; Lin, C.J.;
Rosenfeld, M.G.; Lee, W.H.

#journal

Proc. Natl. Acad. Sci. U.S.A. (1988) 85:6017-6021

#title

Molecular mechanism of retinoblastoma gene inactivation in
retinoblastoma cell line Y79.

#cross-references

MUID:88320373

#contents

annotation

REFERENCE

A38988

#authors

Lees, J.A.; Buchkovich, K.J.; Marshak, D.R.; Anderson, C.W.;
Harlow, E.

#journal

EMBO J. (1991) 10:4279-4290

```

#title      The retinoblastoma protein is phosphorylated on multiple
#contents   sites by human cdc2.
#GENETICS   annotation; phosphorylation sites
#gene       GDB:RB1
##cross-references GDB:118734; OMIM:180200
#map_position 13q14.3-13q14.3
#introns    46/2; 88/3; 127/2; 167/2; 180/2; 203/1; 240/1; 287/3; 313/3;
350/2; 376/2; 405/3; 444/3; 463/3; 474/2; 500/1; 565/3;
605/2; 654/1; 702/3; 737/3; 775/3; 830/2; 840/3; 888/2;
905/1
#note       mutations in the germline gene predispose to hereditary
retinoblastoma; somatic mutations have been found in a
variety of human cancers including retinoblastoma and
osteosarcoma
#superfamily retinoblastoma-associated protein
#cell cycle control; DNA binding; leucine zipper;
osteosarcoma; phosphoprotein; retinoblastoma; transcription
regulation; tumor suppressor
#region alanine-rich\
#region proline-rich\
#region leucine zipper motif\
#binding_site phosphate (Ser) (covalent) (by cdc2
kinase) #status predicted\
#binding_site phosphate (Thr) (covalent) (by cdc2
kinase) #status predicted
#length 928 #molecular-weight 106158 #checksum 3504

Query Match      70.3%; Score 45; DB 1; Length 928;
Best Local Similarity 50.0%; Pred. NO. 2.17e-01;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db      27 PPEDNPVED 36
      :||||:|
Qy      1 APPEDNPVED 10

RESULT 11
ENTRY
TITLE      F69552      #type complete
            leucyl-tRNA synthetase (leus) homolog - Archaeoglobus
            fulgidus
ORGANISM    05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
DATE        21-Aug-1998
ACCESSIONS  F69552
REFERENCE    A69250
#authors     Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
            K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
            Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham,
            D.E.; Kyriakis, N.C.; Fleischmann, R.D.; Quackenbush, J.;
            Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
            Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.;
            Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.;
            Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
            J.F.; McDonald, L.; Yeh, P.; White, O.; Liew, J.; Miller,
            T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;
            D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
            Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
            C.R.; Venter, J.C.
#journal     Nature (1997) 390:364-370
#title       The complete genome sequence of the hyperthermophilic,
            sulfate-reducing archaeon Archaeoglobus fulgidus.
##cross-references MUID:98049343
#accession   F69552
#status      preliminary; nucleic acid sequence not shown;
            translation not shown

#molecule_type DNA
#residues    1-932 #label KLF
##cross-references GB:AE001108; GB:AE000782; NID:g2689431; PID:g2650668;
            TIGR:AR2421
CLASSIFICATION #superfamily valine-tRNA ligase
SUMMARY        #length 932 #molecular-weight 108627 #checksum 6668

```

```

Query Match      70.3%; Score 45; DB 2; Length 932;
Best Local Similarity .66.7%; Pred. NO. 2.17e-01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db      180 PHDONPVED 188
      |::|:||||
Qy      2 PPEDNPVED 10

RESULT 12
ENTRY
TITLE      B55282      #type complete
            neurofibromatosis-related protein NF1 - human
            Gpase activating protein homolog NF1; neurofibromin
            #formal_name Homo sapiens #common_name man
            ORGANISM    10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change
            DATE        20-Mar-1998
ACCESSIONS  B55282; A55282; A35879; A35605; A35910; A35222; A36297;
            I58356
REFERENCE    A55282
#authors     Marchuk, D.A.; Saulino, A.M.; Tavakkol, R.; Swaroop, M.;
            Wallace, M.R.; Andersen, L.B.; Mitchell, A.L.; Gutmann,
            D.H.; Boguski, M.; Collins, F.S.
#journal     Genomics (1991) 11:931-940
#title       cDNA cloning of the type 1 neurofibromatosis gene: complete
            sequence of the NF1 gene product.
##cross-references MUID:92147138
#accession   B55282
#status      not compared with conceptual translation
#molecule_type mRNA
#residues    1-2818 #label MAR
##cross-references GB:M82814; NID:g189164; PID:g189165
#note        sequence extracted from NCBI backbone (NCBIP:80176)
#accession   A55282
#status      preliminary
#molecule_type mRNA
#residues    1-334 #label MA2
#note        sequence extracted from NCBI backbone (NCBIN:80169,
            NCBIP:80172)

REFERENCE    A35879
#authors     Xu, G.; O'Connell, P.; Viskochil, D.; Cawthon, R.; Robertson,
            M.; Culver, M.; Dunn, D.; Stevens, J.; Gesteland, R.;
            White, R.; Weiss, R.
#journal     Cell (1990) 62:599-608
#title       The neurofibromatosis type 1 gene encodes a protein related
            to GAP.
##cross-references MUID:90335969
#accession   A35879
#status      preliminary
#molecule_type mRNA
#residues    335-495, 'I', 497-1555, 'H', 1556-2818 #label XUA
##cross-references GB:M38106; GB:M57449; NID:g189169; PID:g189170
            A35605
REFERENCE    A35605
#authors     Cawthon, R.M.; Weiss, R.; Xu, G.; Viskochil, D.; Culver, M.;
            Stevens, J.; Robertson, M.; Dunn, D.; Gesteland, R.;
            O'Connell, P.; White, R.
#journal     Cell (1990) 62:193-201
#title       A major segment of the neurofibromatosis type 1 gene: cDNA
            sequence, genomic structure, and point mutations.
##cross-references MUID:90304909
#accession   A35605
#status      preliminary
#molecule_type mRNA
#residues    1585-2687 #label CAW
##cross-references EMBL:M38107; EMBL:M57449
            A35910
REFERENCE    A35910
#authors     Cawthon, R.M.; Weiss, R.; Xu, G.; Viskochil, D.; Culver, M.;
            Stevens, J.; Robertson, M.; Dunn, D.; Gesteland, R.;
            O'Connell, P.; White, R.
#journal     Cell (1990) 62:608b
#accession   A35910
#status      preliminary; nucleic acid sequence not shown; not
            compared with conceptual translation

```

```

##molecule_type mRNA
##residues 2688-2818 ##label CA2
REFERENCE A35222
#authors Wallace, M.R.; Marchuk, D.A.; Andersen, L.B.; Letcher, R.;
Odeh, H.M.; Saulino, A.M.; Fountain, J.W.; Brereton, A.;
Nicholson, J.; Mitchell, A.L.; Brownstein, B.H.; Collins,
F.S.
#journal Science (1990) 249:181-186
#title Type 1 neurofibromatosis gene: identification of a large
transcript disrupted in three NF1 patients.
#cross-references MIM:9031972
#accession A35222
##status preliminary
##molecule_type mRNA
##residues 2209-2818 ##label WAL
##cross-references GB:M60496; NID:g189157; PID:g189158; GB:M49193
REFERENCE A36297
#authors Martin, G.A.; Viskochil, D.; Bollag, G.; McCabe, P.C.;
Crosier, W.J.; Haubruck, H.; Conroy, L.; Clark, R.;
O'Connell, P.; Cawthon, R.M.; Innis, M.A.; McCormick, F.
#journal Cell (1990) 63:843-849
#title The GAP-related domain of the neurofibromatosis type 1 gene
product interacts with ras p21.
#cross-references MIM:91029515
#accession A36297
##molecule_type mRNA
##residues 1096-1569 'TPPEPET' ##label MA3
##cross-references GB:M61213; NID:g189162; PID:g189163
##note this clone includes an epitope tag at the 3' end
encoding the sequence TPPEPET, not part of dystrophin
but recognized by the monoclonal antibody K73
REFERENCE I58356
#authors Nishi, T.; Lee, P.S.; Oka, K.; Levin, V.A.; Tanase, S.;
Morino, Y.; Saya, H.
#journal Oncogene (1991) 6:1555-1559
#title Differential expression of two types of the neurofibromatosis
type 1 (NF1) gene transcripts related to neuronal
differentiation.
#cross-references MIM:92019823
#accession I58356
##status translated from GB/EMBL/DBJ
##molecule_type mRNA
##residues 1168-1545 ##label RES
##cross-references GB:M60915; NID:g189159; PID:g189160
GENETICS
#gene GDB:NFI
#map_position 17q11.2-17q11.2
#introns 1370/3
#note the list of introns is incomplete
CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology
KEYWORDS alternative splicing; tumor suppressor
FEATURE
1235-1449 #domain ras-specific GAP catalytic domain homology
#label GAP
#length 2818 #molecular-weight 317030 #checksum 2858
SUMMARY
Query Match 70.3%; Score 45; DB 2; Length 2818;
Best Local Similarity 60.0%; Pred. No. 2.17e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 1526 GPPHKPVAD 1535
:||||:|
QY 1 APPEDNPVED 10

RESULT 13
ENTRY JC5196 #type complete
TITLE neurofibromin-1 - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 20-Feb-1997 #sequence_revision 27-Feb-1997 #text_change
10-Sep-1997
ACCESSIONS JC5196
#status preliminary
#molecule_type mRNA
#residues 1-2825 ##label SUZ
#cross-references DDBJ:D45201; NID:g1841313; PID:d1008732; PID:g1841314
COMMENT This protein contains a GTPase-activating protein-related domain
which is responsible for the stimulatory effect of neurofibromin
on the tyrosinase promoter activity.
CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology
FEATURE
1177-1436 #domain GTPase-activating protein related #status
predicted #label GRD
1237-1451 #domain ras-specific GAP catalytic domain homology
#label GAP
SUMMARY #length 2820 #molecular-weight 317080 #checksum 5628
Query Match 70.3%; Score 45; DB 2; Length 2820;
Best Local Similarity 60.0%; Pred. No. 2.17e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 1528 GPPHKPVAD 1537
:||||:|
QY 1 APPEDNPVED 10

RESULT 14
ENTRY I54352 #type fragment
TITLE neurofibromin - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change
25-Apr-1997
ACCESSIONS I54352
REFERENCE I54352
#authors Bernards, A.; Snijders, A.J.; Hannigan, G.E.; Murthy, A.E.;
Gusella, J.F.
#journal Hum. Mol. Genet. (1993) 2:645-650
#title Mouse neurofibromatosis type 1 cDNA sequence reveals high
degree of conservation of both coding and non-coding mRNA
segments.
#cross-references MIM:93357730
#accession I54352
##status preliminary; translated from GB/EMBL/DBJ
##molecule_type mRNA
#residues 1-2825 ##label RES
#cross-references GB:L10370; NID:g309452; PID:g309453
GENETICS
#gene NFI
CLASSIFICATION #superfamily ras-specific GAP catalytic domain homology
FEATURE
1221-1456 #domain ras-specific GAP catalytic domain homology
#label GAP
#length 2825 #checksum 6076
SUMMARY
Query Match 70.3%; Score 45; DB 2; Length 2825;
Best Local Similarity 60.0%; Pred. No. 2.17e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 1533 GPPHKPVAD 1542
:||||:|
QY 1 APPEDNPVED 10

RESULT 15
ENTRY A28578 #type complete
TITLE pancreatic hormone - ostrich
ALTERNATE_NAMES pancreatic polypeptide
ORGANISM #formal_name Struthio camelus #common_name ostrich

```

DATE 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change
12-Apr-1995
ACCESSIONS A28578
REFERENCE A28578
#authors Litthauer, D.; Oelofsen, W.
#journal Int. J. Pept. Protein Res. (1987) 29:739-745
#title Purification and primary structure of ostrich pancreatic
polypeptide.
#cross-references MUID:87307111
#accession A28578
#molecule_type protein
#residues 1-36 #label LIT
#note the sequence of residues 22-23 was reported as Asn-Asp
in Fig. 7 and as Asp-Asn in Fig. 8
CLASSIFICATION #superfamily pancreatic hormone
SUMMARY #length 36 #molecular-weight 4209 #checksum 2998
Query Match 68.8%; Score 44; DB 2; Length 36;
Best Local Similarity 66.7%; Pred. NO. 3.37e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 8 PGDDAPVED 16
| : | | | |
QY 2 PPEDNPVED 10

Search completed: Thu May 13 15:28:28 1999
Job time : 18 secs.


```

##molecule_type DNA
##residues 1-232 ##label DAV
##cross-references GB:M75136; NID:g331209; PID:g331218
REFERENCE A39447
#authors Davison, A.J.
#journal Virology (1992) 186:9-14
#title Channel catfish virus: a new type of herpesvirus.
#cross-references MUID:92087490
#contents annotation
#note neither amino acid nor nucleotide sequence is given
GENETICS
#gene
CLASSIFICATION
#superfamily Ictalurid herpesvirus 25.5K membrane protein
transmembrane protein
FEATURE
166-182
195-211
#domain transmembrane #status predicted #label TM1\
#domain transmembrane #status predicted #label TM2
#length 232 #molecular-weight 25466 #checksum 8650
SUMMARY
Query Match 74.6%; Score 50; DB 1; Length 232;
Best Local Similarity 50.0%; Pred. No. 1.10e+00;
Matches 5; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
Db 88 ESPGEPHRD 97
:|||||:
QY 1 DGPTGEPOQE 10

RESULT 3
ENTRY #type complete
TITLE hypothetical protein sll1601 - Synchocystis sp. (strain PCC
6803)
ORGANISM #formal_name Synchocystis sp.
#variety PCC 6803
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998
ACCESSIONS S75052
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular
cyanobacterium Synchocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MUID:97061201
#accession S75052
#status preliminary
#molecule_type DNA
#residues 1-414 ##label KAN
#cross-references EMBL:D90910; GB:AB001339; NID:g1652956; PID:d1018647;
PID:g1652997
#note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
SUMMARY #length 414 #molecular-weight 44810 #checksum 1664

Query Match 71.6%; Score 48; DB 2; Length 414;
Best Local Similarity 60.0%; Pred. No. 2.92e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 318 EGPTEAAQQQ 327
:|||||:
QY 1 DGPTGEPOQE 10

RESULT 4
ENTRY #type complete
TITLE tim (timeless) protein - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster

```

```

DATE 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change
16-Feb-1997
ACCESSIONS A57655
REFERENCE A57655
#authors Myers, M.P.; Wager-Smith, K.; Wesley, C.S.; Young, M.W.;
Sehgal, A.
#journal Science (1995) 270:805-808
#title Positional cloning and sequence analysis of the Drosophila
clock gene, timeless.
#accession A57655
#status preliminary; nucleic acid sequence not shown
#molecule_type DNA
#residues 1-1388 ##label MYE
#cross-references GB:U37018
GENETICS
#gene tim
#cross-references FlyBase:FBgn0014396
SUMMARY #length 1388 #molecular-weight 155537 #checksum 3459

Query Match 70.1%; Score 47; DB 2; Length 1388;
Best Local Similarity 60.0%; Pred. No. 4.72e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 510 DGPGKPPQHQ 519
:|||||:
QY 1 DGPTGEPOQE 10

RESULT 5
ENTRY #type complete
TITLE 6-deoxyerythronolide B synthase II - Saccharopolyspora
erythraea
ORGANISM #formal_name Saccharopolyspora erythraea
DATE 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change
07-Aug-1998
ACCESSIONS S23070; S22011
REFERENCE S23070; S22011
#authors Bevilitt, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.
Eur. J. Biochem. (1992) 204:39-49
#journal 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora
erythraea. Cloning of the structural gene, sequence
analysis and inferred domain structure of the
multifunctional enzyme.
#accession S23070
#molecule_type DNA
#residues 1-3573 ##label BEV1
#cross-references EMBL:X62569
REFERENCE S22011
#authors Bevilitt, D.J.
#submission submitted to the EMBL Data Library, September 1991
#accession S22011
#molecule_type DNA
#residues 1-184, 'I', 186-301, 'S', 303-521, 523-658, 'A', 660-993,
1001-1212, 'H', 1214-1392, 1394-2481, 'V', 2482-2827, 'P',
2829-2833, 'L', 2835-2856, 2858-2907, 'A', 2908-3135, 'K',
3137-3166, 'H', 3168-3176, 'L', 3177-3479, 'DH', 3480-3572
##label BEV2
#cross-references EMBL:X62569; NID:g46977; PID:g581651
GENETICS
#gene eryA
#start_codon GTG
CLASSIFICATION
#superfamily [acyl-carrier-protein] S-malonyltransferase
homology; 3-oxoacyl-[acyl-carrier-protein] synthase I
homology; acyl carrier protein homology; long-chain alcohol
dehydrogenase homology; short-chain alcohol dehydrogenase
homology
multifunctional enzyme
KEYWORDS
FEATURE
52-453
#domain 3-oxoacyl-[acyl-carrier-protein] synthase I
homology #label OAS1\
561-843 #domain [acyl-carrier-protein] S-malonyltransferase
homology #label AMT1\
1140-1308 #domain short-chain alcohol dehydrogenase homology

```

```
#label SAD2\
#domain acyl carrier protein homology #label ACPI\
#domain 3-oxoacyl-[acyl-carrier-protein] synthase I
1519-1919 homology #label OAS2\
2023-2305 #domain [acyl-carrier-protein] S-malonyltransferase
2848-3131 homology #label AMT2\
3149-3327 #domain long-chain alcohol dehydrogenase homology #label
LADH\
#domain short-chain alcohol dehydrogenase homology
3420-3493 #domain acyl carrier protein homology #label ACP2
SUMMARY #length 3573 #molecular-weight 375258 #checksum 9583

Query Match 70.1%; Score 47; DB 2; Length 3573;
Best Local Similarity 87.5%; Pred. No. 4.72e+00;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1384 GPTGEPQAQ 1391
QY 2 GPTGEPQQ 9

RESULT 6
ENTRY #type fragment
TITLE collagen alpha 1(I) chain precursor - baboon (fragment)
ORGANISM #formal_name Papio sp. #common_name baboon
DATE 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change
21-Jul-1995
ACCESSIONS A05249
REFERENCE A92078
#authors Epstein Jr., E.H.; Scott, R.D.; Miller, E.J.; Piez, K.A.
#journal J. Biol. Chem. (1971) 246:1718-1724
#title Isolation and characterization of the peptides derived from
soluble human and baboon skin collagen after cyanogen
bromide cleavage.
#cross-references MUID:71134791
#contents CNBr0-1, CNBr2, CNBr4, CNBr5, composition
#accession A05249
#molecule_type protein
#residues 1-140 #label EPS
#experimental_source skin
COMMENT Prolines at the third position of the tripeptide repeating unit
(G-X-Y) are hydroxylated in some or all of the chains.
CLASSIFICATION #superfamily collagen alpha 1(I) chain; fibrillar collagen
carboxyl-terminal homology; von Willebrand factor type C
repeat homology
KEYWORDS coiled coil; extracellular matrix; glycoprotein; skin;
trimer; triple helix
SUMMARY #length 140 #checksum 7697

Query Match 67.2%; Score 45; DB 2; Length 140;
Best Local Similarity 44.4%; Pred. No. 1.21e+01;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 41 ZGPPGZPGZ 49
QY 1 DGPTGEPQQ 9

RESULT 7
ENTRY #type complete
TITLE Probable membrane protein ybcB - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
13-Sep-1998
ACCESSIONS G64831
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
```

```
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession G64831
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-259 #label BLAT
#cross-references GB:AE000194; GB:U00096; NID:g1787148; PID:g1787150;
UMGP:b0920
#experimental_source strain K-12, substrain MG1655
GENETICS
#gene ybcB
KEYWORDS transmembrane protein
FEATURES
13-29 #domain transmembrane #status predicted #label TM1\
39-55 #domain transmembrane #status predicted #label TM2\
SUMMARY #length 259 #molecular-weight 28666 #checksum 3266

Query Match 67.2%; Score 45; DB 2; Length 259;
Best Local Similarity 66.7%; Pred. No. 1.21e+01;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 251 GSSGEPQOE 259
QY 2 GPTGEPQOE 10

RESULT 8
ENTRY #type complete
TITLE MHC class II histocompatibility antigen B-LBII - chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 03-Jun-1993 #sequence_revision 30-Sep-1993 #text_change
08-Sep-1997
ACCESSIONS A45838
REFERENCE A45838
#authors Zoorob, R.; Behar, G.; Kroemer, G.; Auffray, C.
#journal Immunogenetics (1990) 31:179-187
#title Organization of a functional chicken class II B gene.
#accession A45838
#status preliminary
#molecule_type DNA
#residues 1-263 #label ZOO
#cross-references GB:M29763; NID:g212297; PID:g212298
#note the authors translated the codon GTG for residue 112 as
Asp
CLASSIFICATION #superfamily class II histocompatibility antigen;
immunoglobulin homology
SUMMARY #length 263 #molecular-weight 29112 #checksum 5224

Query Match 67.2%; Score 45; DB 2; Length 263;
Best Local Similarity 70.0%; Pred. No. 1.21e+01;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 76 DPLGEPQAE 85
QY 1 DGPTGEPQOE 10

RESULT 9
ENTRY #type fragment
TITLE P53-binding protein 1 - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 09-Mar-1996 #sequence_revision 09-Mar-1996 #text_change
09-Mar-1996
ACCESSIONS I38604
REFERENCE I38604
#authors Iwabuchi, K.; Bartel, P.L.; Li, B.; Marraccino, R.; Fields,
S.
#journal Proc. Natl. Acad. Sci. U.S.A. (1994) 91:6098-6102
#title Two cellular proteins that bind to wild-type but not mutant
p53.
#cross-references MUID:94286584
#accession I38604
#status preliminary; nucleic acid sequence not shown
```

```

##molecule_type mRNA
##residues 1-1027 ##label RES
##cross-references EMBL:U09477; NID:g488591; PID:g488592
SUMMARY
#length 1027 #checksum 8683

Query Match 67.2%; Score 45; DB 2; Length 1027;
Best Local Similarity 60.0%; Pred. No. 1.21e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 808 DPGTGSSEE 817
|||||:::|
Qy 1 DPGTGPQOE 10

RESULT 10
ENTRY CGHU6C #type complete
TITLE Collagen alpha 1(II) chain precursor - human
ALTERNATE_NAMES collagen alpha 1(II) chain
CONTAINS chondrocalcin; collagen alpha 1(II) chain precursor splice
form 1; collagen alpha 1(II) chain precursor splice form 2;
collagen alpha 3(XI) chain
ORGANISM #formal_name Homo sapiens #common_name man
DATE 28-May-1986 #sequence_revision 01-Sep-1995 #text_change
02-Sep-1997
ACCESSIONS
A38513; S06715; S24270; A24828; S06496; A35428; A30147;
A33116; S63514; S04892; S05000; A44309; S16502; A02858;
A27280; A57033; A21733; B21733; A24561; S59491; I84433;
I37250; I37251; I37252; I37253; I37254; I55338; I59535;
I61910
REFERENCE
#authors Ryan, M.C.; Sieraski, M.; Sandell, L.J.
#journal Genomics (1990) 8:41-48
#title The human type II procollagen gene: identification of an
additional protein-coding domain and location of potential
regulatory sequences in the promoter and first intron.
#cross-references MUID:91184811
#accession A38513
##molecule_type DNA
##residues 1-103 ##label RYA
##cross-references GB:M60299; NID:g180883; PID:g18088
REFERENCE
#authors Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.
#journal Nucleic Acids Res. (1989) 17:9473
#title Nucleotide sequence of the full length cDNA encoding for
human type II procollagen.
#cross-references MUID:90067946
#accession S06715
##molecule_type mRNA
##residues 1-28,'R',99-1487 ##label SU2
##cross-references EMBL:X16468; NID:g29515; PID:g29516
#note alternative splice form 1
REFERENCE
#authors Vikkula, M.; Metsaeranta, M.; Syvaenen, A.C.; Ala-Kokko, L.;
Vuorio, E.; Peltonen, L.
#journal Biochem. J. (1992) 285:287-294
#title Structural analysis of the regulatory elements of the type-II
procollagen gene. Conservation of promoter and first intron
sequences between human and mouse.
#cross-references MUID:92344585
#accession S24270
##status translation not shown
##molecule_type DNA
##residues 1-28 ##label VIK
##cross-references EMBL:X58709; GB:S40537; NID:g35659
REFERENCE
#authors Nunez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.
#journal Gene (1986) 44:11-16
#title Promoter region of the human pro-alpha-1-(II)-collagen gene.
#cross-references MUID:87031574
#accession A24828
##molecule_type DNA
##residues 1-8,'T',10-28 ##label NUN
##cross-references GB:M25698; NID:g180872; PID:g553237

REFERENCE
#authors Baldwin, C.T.; Reginato, A.M.; Smith, C.; Jimenez, S.A.;
Prockop, D.J.
#journal Biochem. J. (1989) 262:521-528
#title Structure of cDNA clones coding for human type II
procollagen. The alpha-1(II) chain is more similar to the
alpha-1(I) chain than two other alpha chains of fibrillar
collagens
#cross-references MUID:90026318
#accession S06496
##molecule_type mRNA
##residues 7-28,'R',99-157,'P',159-440,'G',442-456,'E',458-640,'A',
642-831,'PA',834,'P',836-1005,'K',1007-1036,'Q',
1038-1229 ##label BAL
##cross-references EMBL:X16711; NID:g30040; PID:g30041
#note alternative splice form 1
REFERENCE
#authors Ryan, M.C.; Sandell, L.J.
#journal J' Biol. Chem. (1990) 265:10334-10339
#title Differential expression of a cysteine-rich domain in the
amino-terminal propeptide of type II (cartilage)
procollagen by alternative splicing of mRNA.
#cross-references MUID:90285153
#accession A35428
##status not compared with conceptual translation
##molecule_type mRNA
##residues 27-81,'L',83-103 ##label RYA2
#note alternative splice form 2; splicing appears to be under
developmental regulation
REFERENCE
#authors Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.
#journal Genomics (1989) 4:438-441
#title Organization of the exons coding for Pro alpha-1(II) collagen
N-propeptide confirms a distinct evolutionary history of
this domain of the fibrillar collagen genes.
#cross-references MUID:89233138
#accession A30147
##molecule_type DNA
##residues 104-157,'P',159-236 ##label SUM
##cross-references GB:J03065; GB:M23660; GB:M25655; GB:M25656;
GB:M25730; GB:M32168; GB:M64345; NID:g180867;
PID:g180869
REFERENCE
#authors Ala-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.
#journal Proc. Natl. Acad. Sci. U.S.A. (1990) 87:6565-6568
#title Single base mutation in the type II procollagen gene (COL2A1)
as a cause of primary osteoarthritis associated with a mild
chondrodysplasia.
#cross-references MUID:90370826
#accession A33116
##molecule_type DNA
##residues 171-172,'C',174-175 ##label ALA
#note mutant sequence from a family with family with primary
generalized osteoarthritis and mild chondrodysplasia
REFERENCE
#authors Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene
de la Porte, P.; Herbage, D.
#journal Eur. J. Biochem. (1995) 234:125-131
#title Immunohistochemical and biochemical analyses of
20000-25000-year-old fossil cartilage.
#accession S63514
##molecule_type protein
##residues 243-261;575-590;756-763,'X',765-779 ##label FRA
REFERENCE
#authors Ramirez, F.
#journal Submitted to the EMBL Data Library, December 1988
#accession S04892
##molecule_type mRNA
##residues 501-676,'A',678-783,'A',785-831,'PA',834,'F',836-1214
#note #label RAM
#cross-references EMBL:X13783; NID:g30037; PID:g930050
REFERENCE
#authors Vikkula, M.; Peltonen, L.

```

```

#journal FBS Lett. (1989) 250:171-174
#title Structural analyses of the polymorphic area in type II
#cross-references MUID:89325561
#accession S05000
##molecule_type DNA
##residues 630-640,'A',642-785 #label VIK2
##cross-references EMBL:X16158; NID:g29951
REFERENCE
A44309
#authors Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin,
D.L.; Cohn, D.H.; Eyre, D.R.
#journal J. Biol. Chem. (1992) 267:22522-22526
#title An amino acid substitution (Gly853-->Glu) in the collagen
alpha 1(I) chain produces hypochondrogenesis.
#cross-references MUID:9305458
#accession A44309
##status nucleic acid sequence not shown; not compared with
conceptual translation
##molecule_type DNA; mRNA
##residues 752-831,'PA',834,'F',836-1005,'K',1007-1036,'Q',
1038-1052,'E',1054-1068,'T',1070-1097 #label BOG
##cross-references GB:L00977; NID:g180812
##note Sequence extracted from NCBI backbone (NCBIP:117273);
parts of this sequence were determined by protein
sequencing
##note mutant sequence associated with perinatal lethal
hypochondrogenesis
REFERENCE
S16502
#authors Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
#journal Proc. Natl. Acad. Sci. U.S.A. (1990) 87:3889-3893
#title Tandem duplication within a type II collagen gene (COL2A1)
exon in an individual with spondyloepiphyseal dysplasia.
#cross-references MUID:90251662
#accession S16502
##molecule_type DNA
##residues 1164-1184,'GPSGKGANGIPGP',1185-1199 #label TIL
##cross-references EMBL:M37126; NID:g180808; PID:g180809
##note mutant sequence from a patient with spondyloepiphyseal
dysplasia
REFERENCE
A02858
#authors Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.;
Solomon, E.
#journal Proc. Natl. Acad. Sci. U.S.A. (1985) 82:2555-2559
#title Identification and characterization of the human type II
collagen gene (COL2A1).
#cross-references MUID:85190534
#accession A02858
##molecule_type DNA
##residues 1032-1056,'N',1058-1068,'T',1070-1487 #label CHE
##cross-references GB:J00116; NID:g180395; PID:g180396
REFERENCE
A27280
#authors Elima, K.; Vuorio, T.; Vuorio, E.
#journal Nucleic Acids Res. (1987) 15:9499-9504
#title Determination of the single polyadenylation site of the human
pro-alpha-1(I) collagen gene.
#cross-references MUID:88067771
#accession A27280
##molecule_type DNA; mRNA
##residues 1175-1487 #label ELI
##cross-references EMBL:X06268; NID:g30096; PID:g30097
##experimental_source fetal epiphyseal cartilage
REFERENCE
A57033
#authors van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
#journal Biochem. J. (1986) 237:923-925
#title Chondrocalcin is identical with the C-propeptide of type II
procollagen.
#accession A57033
##molecule_type protein
##residues 'XE',1244-1246,'N',1248,'X',1250-1265;1295-1305;
...
Note: remainder of annotations omitted.

```

Query Match 67.2%; Score 45; DB 1; Length 1487;

Best Local Similarity 66.7%; Pred. No. 1.21e+01;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 999 GPSGEPGQQ 1007
Qy 2 GPTGEPQOE 10
||:|||||:

RESULT 11
ENTRY A33106 #type complete
TITLE neurogenic locus mam protein - fruit fly (Drosophila melanogaster)
ALTERNATE_NAMES mastermind protein
ORGANISM #formal_name Drosophila melanogaster
DATE 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 24-Sep-1998

ACCESSIONS A36391; A33106; S13514
REFERENCE A36391
#authors Smoller, D.; Friedel, C.; Schmid, A.; Bettler, D.; Lam, L.; Yedvobnick, B.
#journal Genes Dev. (1990) 4:1688-1700
#title The Drosophila neurogenic locus mastermind encodes a nuclear protein unusually rich in amino acid homopolymers.
#cross-references MUID:91065516
#accession A36391

GENETICS
#status preliminary
##molecule_type mRNA
##residues 1-1596 #label SMO
##cross-references GB:X54251; NID:g8203; PID:g8204
##note strain Canton S

#gene FlyBase:mam
#cross-references FlyBase:FBgn0002643
SUMMARY #length 1596 #molecular_weight 167717 #checksum 4406

Query Match 67.2%; Score 45; DB 2; Length 1596;
Best Local Similarity 66.7%; Pred. No. 1.21e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 1317 GPMGPGQQO 1325
Qy 2 GPTGEPQOE 10
||:|||||:

RESULT 12
ENTRY I54523 #type fragment
TITLE kinesin-related protein - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 05-Jun-1998

ACCESSIONS I54523
REFERENCE I54523
#authors Ando, A.; Yara-Kikutl, Y.; Kawata, H.; Okamoto, N.; Imai, T.; Eki, T.; Yokoyama, K.; Soeda, E.; Ikemura, T.; Abe, K.; Inoko, H.

#journal Immunogenetics (1994) 39:194-200
#title Cloning of a new kinesin-related gene located at the centromeric end of the human MHC region.

#cross-references MUID:94102819
#accession I54523
#status preliminary; translated from GB/EMBL/DBJ

##molecule_type mRNA
##residues 1-519 #label RES

##cross-references GB:D14678; NID:g510281; PID:g510282
CLASSIFICATION #superfamily unassigned kinesin-related proteins; kinesin motor domain homology

ATP; P-loop

KEYWORDS
FEATURE
157-515 #domain kinesin motor domain homology #label KMOT\
256-263 #region nucleotide-binding motif A (P-loop)
SUMMARY #length 519 #checksum 4377

Query Match 65.7%; Score 44; DB 2; Length 519;

```

Best Local Similarity 66.7%; Pred. No. 1.92e+01;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 269 GPGGDPQLE 277
   |||:|||||
QY 2 GPTGEPQOE 10

RESULT 13
ENTRY S30059 #type complete
TITLE Probable transport protein ALD - human
ALTERNATE_NAMES
ORGANISM adrenoleukodystrophy-related protein
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
24-Sep-1998
ACCESSIONS S30059
REFERENCE Mosser, J.; Douar, A.M.; Sarde, C.O.; Kioschis, P.; Feil, R.;
#authors Moser, H.; Poustka, A.M.; Mandel, J.L.; Aubourg, P.
#journal Nature (1993) 361:726-730
#title Putative X-linked adrenoleukodystrophy gene shares unexpected
#accession homology with ABC transporters.
#molecule_type mRNA
#residues 1-745 #label MOS
#cross-references EMBL:221876; NID:g38590; PID:g38591
GENETICS
#gene GDB:ALD
#cross-references GDB:118991; OMIM:300100
#map_position Xq28-Xq28
CLASSIFICATION #superfamily ATP-binding cassette homology
KEYWORDS adrenoleukodystrophy; ATP; membrane protein; P-loop
FEATURE
490-675 #domain ATP-binding cassette homology #label ABC\
507-514 #region nucleotide-binding motif A (P-loop)\
513 #binding_site ATP (lys) #status predicted
SUMMARY #length 745 #molecular-weight 82908 #checksum 9250

Query Match 65.7%; Score 44; DB 2; Length 745;
Best Local Similarity 60.0%; Pred. No. 1.92e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 47 QAPAGEPQOE 56
   |||:|||||
QY 1 DGTGEPQOE 10

RESULT 14
ENTRY G02500 #type complete
TITLE adrenoleukodystrophy protein - human
ORGANISM adrenoleukodystrophy-related protein
DATE 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change
18-Sep-1998
ACCESSIONS G02500
REFERENCE H01367
#authors Platzer, M.; Bauer, D.; Drescher, B.
#submission submitted to the EMBL Data Library, March 1996
#accession G02500
#status preliminary; translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues 1-745 #label PIA
#cross-references EMBL:U52111; NID:g1302649; PID:g1302652
GENETICS
#gene GDB:ALD; AMN
#cross-references GDB:118991; OMIM:300100
#map_position Xq28-Xq28
#introns 300/3; 361/1; 408/3; 465/1; 496/3; 545/2; 594/1; 622/2; 664/2
CLASSIFICATION #superfamily ATP-binding cassette homology
FEATURE
490-675 #domain ATP-binding cassette homology #label ABC\
507-514 #region nucleotide-binding motif A (P-loop)
#length 745 #molecular-weight 82936 #checksum 9439

```

```

Query Match 65.7%; Score 44; DB 2; Length 745;
Best Local Similarity 60.0%; Pred. No. 1.92e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 47 QAPAGEPQOE 56
   |||:|||||
QY 1 DGTGEPQOE 10

RESULT 15
ENTRY S14113 #type complete
TITLE 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase (EC
ALTERNATE_NAMES 3.1.4.11) delta-2 - bovine
ORGANISM inositol-phospholipid-specific phospholipase C
DATE 21-Nov-1993 #sequence_revision 02-Jun-1995 #text_change
29-May-1998
ACCESSIONS S14113; S04944
REFERENCE S14113
#authors Meldrum, E.; Kriz, R.W.; Totty, N.; Parker, P.J.
#journal Eur. J. Biochem. (1991) 196:159-165
#title A second gene product of the inositol-phospholipid-specific
#accession phospholipase C-delta subclass.
#cross-references MUID:91160548
#accession S14113
#status preliminary
#molecule_type mRNA
#residues 1-764 #label MEL
#experimental_source brain
REFERENCE S04944
#authors Meldrum, E.; Katan, M.; Parker, P.
#journal Eur. J. Biochem. (1989) 182:673-677
#title A novel inositol-phospholipid-specific phospholipase C. Rapid
#cross-references MUID:89325315
#accession S04944
#status preliminary
#molecule_type protein
#residues 528-541, 'X', 543-553; 659-669 #label ME2
#experimental_source brain
CLASSIFICATION #superfamily 1-phosphatidylinositol-4,5-bisphosphate
phosphodiesterase III; 1-phosphatidylinositol-4,5-bisphosphate
5-bisphosphate phosphodiesterase domain X homology;
1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase
domain Y homology; pleckstrin repeat homology
phosphoric diester hydrolase
KEYWORDS
FEATURE
292-435 #domain 1-phosphatidylinositol-4,5-bisphosphate
phosphodiesterase domain X homology #label PIPX\
489-609 #domain 1-phosphatidylinositol-4,5-bisphosphate
phosphodiesterase domain Y homology #label PIPY
SUMMARY #length 764 #molecular-weight 87681 #checksum 1361

Query Match 65.7%; Score 44; DB 2; Length 764;
Best Local Similarity 85.7%; Pred. No. 1.92e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 340 DGPGSEP 346
   |||:|||||
QY 1 DGTGEP 7

Search completed: Thu May 13 15:36:34 1999
Job time : 12 secs.

```

WATERMANS

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 13 15:45:29 1999; MasPar time 2.00 Seconds
133.991 Million cell updates/sec

Tabular output not generated.

Title: >US-09-040-485-9
Description: (1-10) from US09040485.pep
Perfect Score: 61
Sequence: 1 QENPDSPEPV 10

Scoring table: PAM 150
Gap 15

Searched: 74019 seqs, 26840295 residues

Post-processing: Minimum Match 0\$
Listing first 45 summaries

Database: swiss-prot35
1:swissprot

Statistics: Mean 21.064; Variance 21.843; scale 0.964

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	61	100.0	754	1	ASPH_BOVIN	1.50e-04
2	61	100.0	757	1	ASPH_HUMAN	1.50e-04
3	48	78.7	1379	1	MET_MOUSE	4.64e-01
4	45	73.8	304	1	CBP2_SIMV1	2.51e+00
5	44	72.1	662	1	UL06_HSV6U	4.34e+00
6	43	70.5	277	1	TRT1_HUMAN	7.43e+00
7	43	70.5	355	1	IT44_STRFR	7.43e+00
8	43	70.5	475	1	TRB1_ECOLI	7.43e+00
9	43	70.5	995	1	AGAA_VIBS7	7.43e+00
10	43	70.5	1218	1	YK83_YEAST	7.43e+00
11	43	70.5	1390	1	MET_HUMAN	7.43e+00
12	43	70.5	1816	1	LM44_HUMAN	7.43e+00
13	42	68.9	622	1	SKNT_YEAST	7.43e+00
14	42	68.9	1214	1	TSGA_RAT	1.26e+01
15	41	67.2	244	1	SCS2_YEAST	2.11e+01
16	41	67.2	311	1	SDCL_MOUSE	2.11e+01
17	41	67.2	313	1	SDCL_RAT	2.11e+01
18	41	67.2	539	1	PIR3_CHLTR	2.11e+01
19	41	67.2	642	1	PHSA_STRAT	2.11e+01
20	41	67.2	662	1	UL06_HSV62	2.11e+01
21	41	67.2	727	1	KDGA_RAT	2.11e+01
22	41	67.2	938	1	EBNA_EBV	2.11e+01
23	41	67.2	1857	1	FAS2_PENPA	2.11e+01

RESULT	ID	ASPH_BOVIN	STANDARD;	PRT;	754 AA.
AC	Q28056;				
DT	01-NOV-1997 (REL. 35, CREATED)				
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)				
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)				
DE	ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-HYDROXYLASE)				
DE	GN ASPH.				
OS	BOS TAURUS (BOVINE).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;				
OC	EUTHERIA; ARTIODACTYLA.				
RN	[1]				
RC	SEQUENCE FROM N.A.				
RC	TISSUE=LIVER, AND BRAIN;				
RX	MEDLINE: 92332546.				
RA	JTA S., VANDUSEN W.J., DIEHL R.E., KOHL N.E., DIXON R.A.F.,				
RA	ELLISTON K.O., STERN A.M., FRIEDMAN P.A.;				
RL	J. BIOL. CHEM. 267:14322-14327(1992).				
RL	[2]				
RC	SEQUENCE OF 289-385 AND 615-641.				
RC	TISSUE=LIVER;				
RX	MEDLINE: 91310689.				
RA	WANG O., VANDUSEN W.J., PETROSKI C.J., GARSKY V.M., STERN A.M.,				
RA	FRIEDMAN P.A.;				
CC	FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF PROTEINS.				
CC	-1- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) -				
CC	PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).				
CC	-1- COFACTOR: IRON.				
CC	-1- SUBUNIT: MONOMER.				
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM.				
CC	-1- PTM: MIGHT BE PROCESSED TO THE 56 KD (AA 289-754) OR 52 KD (AA 311-754) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM.				
DR	EMBL: M91213; G162694;				
KW	OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR;				
KW	ENDOPLASMIC RETICULUM.				
FT	DOMAIN 1 57				
FT	TRANSMEM 58 78				
FT	CYTOPLASMIC (POTENTIAL).				
FT	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)				
FT	(POTENTIAL).				
FT	LUMENAL (POTENTIAL).				
FT	POLY-GLY.				
FT	POLY-SER.				

FT DOMAIN 318 328 POLY-LYS.
 FT CARBOHYD 96 96 POTENTIAL.
 FT CARBOHYD 466 466 POTENTIAL.
 FT CARBOHYD 702 702 POTENTIAL.
 SQ SEQUENCE 754 AA; 84998 MW; 608861B2 CRC32;
 Query Match 100.0%; Score 61; DB 1; Length 754;
 Best Local Similarity 100.0%; Pred. No. 1.50e-04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 248 QENPDSSEPV 257
 QY 1 QENPDSSEPV 10
 RESULT 2
 ID ASPH_HUMAN STANDARD; PRT; 757 AA.
 AC Q12797;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE ASPARTYL/ASPARAGINYL BETA-HYDROXYLASE (EC 1.14.11.16) (ASPARTATE BETA-HYDROXYLASE) (ASP BETA-HYDROXYLASE) (PEPTIDE-ASPARTATE BETA-DIOXYGENASE).
 DE ASPH.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; PRIMATES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 95121937.
 RA KORIOTH F., GIEFFERS C., FREY J.;
 RL GENE 150:395-399(1994).
 CC -!- FUNCTION: SPECIFICALLY HYDROXYLATES AN ASP OR ASN RESIDUE IN CERTAIN EPIDERMAL GROWTH FACTOR-LIKE (EGF) DOMAINS OF A NUMBER OF PROTEINS.
 CC -!- CATALYTIC ACTIVITY: PEPTIDE L-ASPARTATE + 2-OXOGLUTARATE + O(2) -> PEPTIDE 3-HYDROXY-L-ASPARTATE + SUCCINATE + CO(2).
 CC -!- COFACTOR: IRON.
 CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC RETICULUM.
 CC -!- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES TESTED.
 CC -!- PTM: MIGHT BE PROCESSED TO THE 56 KD (AA 274-757) OR 52 KD (AA 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY SIMILARITY).
 CC EMBL: U03109; E82591; .
 DR MIM: 600582; .
 DR OXIDOREDUCTASE; DIOXYGENASE; IRON; TRANSMEMBRANE; SIGNAL-ANCHOR; ENDOPLASMIC RETICULUM.
 KW FT DOMAIN 1 54
 FT TRANSMEM 55 75
 FT CYTOPLASMIC (POTENTIAL).
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT DOMAIN 76 757
 FT LUMENAL (POTENTIAL).
 FT POLY-SER.
 FT POLY-LYS.
 FT CARBOHYD 323 332
 FT CARBOHYD 452 452
 FT CARBOHYD 705 705
 FT POTENTIAL.
 SQ SEQUENCE 757 AA; 85498 MW; AE6AFC24 CRC32;
 Query Match 100.0%; Score 61; DB 1; Length 757;
 Best Local Similarity 100.0%; Pred. No. 1.50e-04;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 233 QENPDSSEPV 242
 QY 1 QENPDSSEPV 10
 RESULT 3
 ID MET_MOUSE STANDARD; PRT; 1379 AA.
 AC P16056; Q62125;
 DT 01-APR-1990 (REL. 14, CREATED)

DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE HEPATOCYTE GROWTH FACTOR RECEPTOR PRECURSOR (MET PROTO-ONCOGENE TYROSINE KINASE) (EC 2.7.1.112) (HGF-SF RECEPTOR).
 GN MET.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 88262253.
 RA CHAN A.M.L., KING H.W.S., DEAKIN E.A., TEMPEST P.R., HILKENS J., KROEZEN V., EDWARDS D.R., WILLS A.J., BROOKES P., COOPER C.S.;
 RL ONCOGENE 2:593-599(1988).
 RN [2]
 RP SEQUENCE OF 1199-1270 FROM N.A.
 RX MEDLINE: 90152381.
 RA WILKS A.F., KURBAN R.R., HOVENS C.M., RALPH S.J.;
 RL GENE 85:67-74(1989).
 CC -!- FUNCTION: RECEPTOR FOR HEPATOCYTE GROWTH FACTOR. HAS A TYROSINE-PROTEIN KINASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP + PROTEIN TYROSINE PHOSPHATE.
 CC -!- SUBUNIT: HETERODIMER FORMED OF AN ALPHA CHAIN (50 KD) AND A BETA CHAIN (145 KD) WHICH ARE DISULFIDE LINKED.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- DISEASE: ACTIVATION OF MET AFTER REARRANGEMENT WITH THE TPR (TRANSLOCATED PROMOTER) LOCUS OF CHROMOSOME 1 PRODUCES AN ONCOGENIC PROTEIN.
 CC -!- SIMILARITY: BELONGS TO THE MET TYROSINE KINASE FAMILY OF RECEPTOR.
 CC EMBL: Y0671; G53059; .
 DR EMBL: M33424; G200574; .
 DR PIR: S01254; S01254.
 DR MGI: 96969; MET.
 DR PROSITE: PS00107; PROTEIN KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
 DR TRANSFERASE; TYROSINE-PROTEIN KINASE; PROTO-ONCOGENE; ATP-BINDING; RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; PHOSPHORYLATION; SIGNAL.
 KW FT SIGNAL 1 24
 FT CHAIN 25 1379
 FT HEPATOCYTE GROWTH FACTOR RECEPTOR.
 FT DOMAIN 25 931
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 932 954
 FT POTENTIAL.
 FT DOMAIN 955 1379
 FT CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 1076 1343
 FT PROTEIN KINASE.
 FT SITE 306 307
 FT CLEAVAGE (POTENTIAL).
 FT NP_BIND 1082 1090
 FT ATP (BY SIMILARITY).
 FT BINDING 1108 1108
 FT ATP (BY SIMILARITY).
 FT ACT_SITE 1202 1202
 FT BY SIMILARITY.
 FT MOD_RES 1233 1233
 FT PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 45 45
 FT POTENTIAL.
 FT CARBOHYD 106 106
 FT POTENTIAL.
 FT CARBOHYD 201 201
 FT POTENTIAL.
 FT CARBOHYD 357 357
 FT POTENTIAL.
 FT CARBOHYD 398 398
 FT POTENTIAL.
 FT CARBOHYD 404 404
 FT POTENTIAL.
 FT CARBOHYD 606 606
 FT POTENTIAL.
 FT CARBOHYD 634 634
 FT POTENTIAL.
 FT CARBOHYD 784 784
 FT POTENTIAL.
 FT CARBOHYD 878 878
 FT POTENTIAL.
 FT CONFLICT 1199 1199
 FT V -> I (IN REF. 2).
 FT CONFLICT 1255 1255
 FT T -> R (IN REF. 2).
 FT CONFLICT 1261 1261
 FT K -> T (IN REF. 2).
 FT CONFLICT 1269 1270
 FT VL -> IP (IN REF. 2).
 FT SEQUENCE 1379 AA; 153548 MW; E1597F1A CRC32;
 SQ
 Query Match 78.7%; Score 48; DB 1; Length 1379;
 Best Local Similarity 70.0%; Pred. No. 4.64e-01;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 347 QSKPDSAEVP 356
 QY 1 QENPDSSEPV 10

Best Local Similarity 50.0%; Pred. No. 4.34e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 612 QTHPENSEPI 621
QY 1 QENPDSSEPV 10

RESULT 6

ID TTTL_HUMAN STANDARD; PRT; 277 AA.
AC P13805;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE TROPONIN T, SLOW SKELETAL MUSCLE ISOFORMS.
GN TNNI1 OR TNNI.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88058976.
RA GAHLMANN R., TROUTT A.B., WADE R.P., GUNNING P., KEDES L.;
RL J. BIOL. CHEM. 262:16122-16126(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-SKELETAL MUSCLE;
RX MEDLINE; 94183266.
RA SAMSON F., MESNARD L., MIHOVILOVIC M., POTTER T.G., MERCADIER J.-J.,
RA ROSES A.D., GILBERT J.R.;
RL BIOCHEM. BIOPHYS. RES. COMMUN. 199:841-847(1994).
CC -1- FUNCTION: TROPONIN T IS THE TROPOMYOSIN-BINDING SUBUNIT OF
CC CALCIUM-SENSITIVITY TO STRIATED MUSCLE ACTOMYOSIN ATPASE ACTIVITY.
CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT ISOFORMS ARE GENERATED BY A
CC DIFFERENTIAL RNA SPLICING MECHANISM.
DR EMBL; M19309; G339781; -;
DR EMBL; M19308; G339783; -;
DR EMBL; S69208; G546021; -;
DR EMBL; S69209; G546023; -;
DR PIR; A29783; A29783.
DR MIM; 191041; -;
KW MUSCLE PROTEIN; ALTERNATIVE SPLICING; MULTIGENE FAMILY.
FT INIT_MET 0 0
FT MOD_RES 1 1
FT PHOSPHORYLATION (BY CK2)
FT (BY SIMILARITY).
FT VARSPLIC 24 34
FT MISSING (IN SECOND ISOFORM).
FT VARSPLIC 204 219
FT MISSING (IN SECOND AND THIRD ISOFORMS).
FT CONFLICT 19 19
FT E -> D (IN REF. 1).
SQ SEQUENCE 277 AA; 32817 MW; B0685CC1 CRC32;

Query Match 70.5%; Score 43; DB 1; Length 277;

Best Local Similarity 50.0%; Pred. No. 7.43e+00;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 22 BEAPEPEPV 31
QY 1 QENPDSSEPV 10

RESULT 7

ID YT44_STRFR STANDARD; PRT; 395 AA.
AC P20188;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 44.4 KD PROTEIN IN TRANSPOSON TN4556.
OS STREPTOMYCES FRADIAE.
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-TN4556;
RX MEDLINE; 90185236.

Best Local Similarity 50.0%; Pred. No. 2.51e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 17 QEHPEHVEPV 26
QY 1 QENPDSSEPV 10

RESULT 5

ID UL05_HSV6U STANDARD; PRT; 662 AA.
AC P52453;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE VIRION PROTEIN U76.
GN U76 OR HDLFI.
OS HERPES SIMPLEX VIRUS (TYPE 6 / STRAIN UGANDA-1102).
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; BETAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95027704.
RA NICHOLAS J.;
RL VIROLOGY 204:738-750(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95266321.
RA GOMPELS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,
RA MARTIN M.E., EFSTATHIOU S., CRAXTON M., MACAULAY H.A.;
RL VIROLOGY 209:29-51(1995).
CC -1- FUNCTION: PRESUMED VIRION PROTEIN; POSSIBLE ROLE IN DNA
CC PACKAGING.
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL6,
CC EHV-1 56, EBV BBRF1, HCMV UL104, AND VZV 54.
DR EMBL; U13194; G662099; -;
DR EMBL; X83413; G854055; -;
SQ SEQUENCE 662 AA; 77234 MW; 1F2C2F67 CRC32;

Query Match 73.8%; Score 45; DB 1; Length 304;

Best Local Similarity 50.0%; Pred. No. 2.51e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 17 QEHPEHVEPV 26
QY 1 QENPDSSEPV 10

RESULT 5

ID UL05_HSV6U STANDARD; PRT; 662 AA.
AC P52453;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE VIRION PROTEIN U76.
GN U76 OR HDLFI.
OS HERPES SIMPLEX VIRUS (TYPE 6 / STRAIN UGANDA-1102).
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; BETAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95027704.
RA NICHOLAS J.;
RL VIROLOGY 204:738-750(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95266321.
RA GOMPELS U.A., NICHOLAS J., LAWRENCE G., JONES M., THOMSON B.J.,
RA MARTIN M.E., EFSTATHIOU S., CRAXTON M., MACAULAY H.A.;
RL VIROLOGY 209:29-51(1995).
CC -1- FUNCTION: PRESUMED VIRION PROTEIN; POSSIBLE ROLE IN DNA
CC PACKAGING.
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL6,
CC EHV-1 56, EBV BBRF1, HCMV UL104, AND VZV 54.
DR EMBL; U13194; G662099; -;
DR EMBL; X83413; G854055; -;
SQ SEQUENCE 662 AA; 77234 MW; 1F2C2F67 CRC32;

Query Match

72.1%; Score 44; DB 1; Length 662;

```

RA SIEMENIAK D.R., SLIGHTOM J.L., CHUNG S.T.;
RL GENE 86:1-9(1990).
DR EMBL; M29297; G1196913; -.
DR PIR; J00430; J00430.
KW HYPOTHETICAL PROTEIN; TRANSPOSABLE ELEMENT.
SQ SEQUENCE 395 AA; 44379 MW; 07E0B910 CRC32;

Query Match 70.5%; Score 43; DB 1; Length 395;
Best Local Similarity 60.0%; Pred. No. 7.43e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 58 QERADGAEVP 67
|||:|:|
QY 1 QENPDSSEPV 10

RESULT 8
ID TRB1_ECOLI STANDARD; PRT; 475 AA.
AC P41067;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE TRAB PROTEIN.
GN TRAB.
OS ESCHERICHIA COLI.
OC PLASMID F. AND PLASMID COLB2.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC ENTEROBACTERIACEAE.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE; 94359430.
RX FROST L.S., IPPEN-IHLER K., SKURRAY R.A.;
RL MICROBIOL. REV. 58:162-210(1994).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN-K12; PLASMID-COLB2;
RC MEDLINE; 96236035.
RX ANTHONY K.G., KATHIR P., MOORE D., IPPEN-IHLER K., FROST L.S.;
RL J. BACTERIOL. 178:3194-3200(1996).
CC -1- FUNCTION: INVOLVED IN F PILUS ASSEMBLY.
DR EMBL; U01159; G398503; -.
DR EMBL; U51860; G1293090; -.
KW PLASMID; CONJUGATION.
SQ SEQUENCE 475 AA; 50460 MW; E85059FB CRC32;

Query Match 70.5%; Score 43; DB 1; Length 475;
Best Local Similarity 50.0%; Pred. No. 7.43e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 426 EDNPESPVPV 435
::|:|:|
QY 1 QENPDSSEPV 10

RESULT 9
ID AGAA_VIB7 STANDARD; PRT; 995 AA.
AC P48839;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE BETA-AGARASE A PRECURSOR (EC 3.2.1.81) (AGARASE 0107).
GN AGAA.
OS VIBRIO SP. (STRAIN JT0107).
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC VIBRIONACEAE.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE; 94113702.
RX SUGANO Y., MATSUMOTO T., KODAMA H., NOMA M.;
RL APPL. ENVIRON. MICROBIOL. 59:3750-3756(1993).
CC -1- FUNCTION: HYDROLYSES AGAROSE AND ALSO NEOGAROTETRAOSE TO YIELD
CC NEOGAROBIOSE.
CC CATALYTIC ACTIVITY: HYDROLYSIS OF 1,3-BETA-D-GALACTOSIDIC LINKAGES

```

```

CC -1- SIMILARITY: BELONGS TO THE TETRAMER AS THE PREDOMINANT PRODUCT.
CC EMBL; D14721; G497893; -.
DR HYDROLASE; GLYCOSIDASE; SIGNAL; MULTIGENE FAMILY.
KW SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 995 BETA-AGARASE A.
FT DOMAIN 781 784 POLY-ALA.
FT DOMAIN 971 974 POLY-GLY.
SQ SEQUENCE 995 AA; 107275 MW; 9FA46198 CRC32;

Query Match 70.5%; Score 43; DB 1; Length 995;
Best Local Similarity 66.7%; Pred. No. 7.43e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 947 EAPDGEVPV 955
|:|:|:|
QY 2 ENPDSSEPV 10

RESULT 10
ID YK83_YEAST STANDARD; PRT; 1218 AA.
AC P36028;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE PROBABLE ATP-DEPENDENT PERMEASE YKR103W.
GN YKR103W.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
[1]
RN SEQUENCE FROM N.A.
RP GAILLON L., DUJON B.;
RL SUBMITTED (MAR-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). BELONGS TO THE MDR SUBFAMILY.
DR EMBL; Z28328; G486611; -.
DR PIR; S38182; S38182.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW HYPOTHETICAL PROTEIN; ATP-BINDING; TRANSMEMBRANE; GLYCOPROTEIN;
KW TRANSPORT.
FT TRANSMEM 136 152 POTENTIAL.
FT TRANSMEM 171 195 POTENTIAL.
FT TRANSMEM 303 327 POTENTIAL.
FT TRANSMEM 349 367 POTENTIAL.
FT TRANSMEM 455 473 POTENTIAL.
FT TRANSMEM 480 502 POTENTIAL.
FT TRANSMEM 559 582 POTENTIAL.
FT TRANSMEM 589 605 POTENTIAL.
FT DOMAIN 606 949 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 950 971 POTENTIAL.
FT TRANSMEM 1010 1034 POTENTIAL.
FT TRANSMEM 1082 1103 POTENTIAL.
FT TRANSMEM 1110 1129 POTENTIAL.
FT NP_BIND 686 693 ATP (POTENTIAL).
SQ SEQUENCE 1218 AA; 137996 MW; 424A6C36 CRC32;

Query Match 70.5%; Score 43; DB 1; Length 1218;
Best Local Similarity 77.8%; Pred. No. 7.43e+00;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 416 ENPDSSEA 424
:|:|:|:|
QY 1 QENPDSSEPV 9

RESULT 11
ID MET_HUMAN STANDARD; PRT; 1390 AA.
AC P08581;
DT 01-AUG-1988 (REL. 08, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE HEPATOCYTE GROWTH FACTOR RECEPTOR PRECURSOR (MET PROTO-ONCOGENE

```

DE TYROSINE KINASE (EC 2.7.1.112) (HGF-SF RECEPTOR).
GN MET.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RA GIORDANO S.;
RL SUBMITTED (NOV-1990) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87317655.
RA PARK M., DEAN M., KAUL K., BRAUN M.J., GONDA M.A., VANDE WOUDE G.;
RL PROC. NATL. ACAD. SCI. U.S.A. 84:6379-6383(1987).
RN [3]
RP SEQUENCE OF 1010-1390 FROM N.A.
RX MEDLINE; 88143699.
RA CHAN A.M.L., KING H.W.S., TEMPEST P.R., DEAKIN E.A., COOPER C.S.,
RA BROOKES P.;
RL ONCOGENE 1:229-233(1987).
RN [4]
RP SEQUENCE OF 1206-1264 FROM N.A.
RX MEDLINE; 94067791.
RA LEE S.T., STRUNK K.M., SPRITZ R.A.;
RL ONCOGENE 8:3403-3410(1993).
RN [5]
RP SEQUENCE OF 1267-1390 FROM N.A.
RX MEDLINE; 86063462.
RA DEAN M., PARK M., LE BEAU M.M., ROBINS T.S., DIAZ M.O., ROWLEY J.D.,
RA BLAIR D.G., VANDE WOUDE G.F.;
RL NATURE 318:385-388(1985).
RN [6]
RP SEQUENCE OF 1-754 FROM N.A.
RA PAULEY A., ANDREWS S.;
RL SUBMITTED (MAY-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [7]
RP FUNCTION.
RX MEDLINE; 911118019.
RA BOTTARO D.P., RUBIN J.S., FALETTO D.L., CHAN A.M.-L., KWIECIK T.E.,
RA VANDE WOUDE G.F., AARONSON S.A.;
RL SCIENCE 251:802-804(1991).
RN [8]
RP PHOSPHORYLATION AT TYR-1235.
RX MEDLINE; 92011756.
RA FERRACINI R., LONGATI P., NALDINI L., VIGNA E., COMOGGIO P.M.;
RL J. BIOL. CHEM. 266:19558-19564(1991).
RN [9]
RP VARIANT'S HPRC, AND VARIANT VAL-320.
RX MEDLINE; 97285124.
RA SCHMIDT L., DUH F.-M., CHEN F., KISHIDA T., GLENN G., CHOYKE P.,
RA SCHERER S.W., ZHUANG Z., LUBENSKY I., DEAN M., ALLIKMETS R.,
RA CHIDAMBARAM A., BERGERHEIM U.R., FELTIS J.T., CASADEVALL C.,
RA ZAMARRON A., BERNUES M., RICHARD S., LIPS C.J.M., WALTHER M.M.,
RA TSUI L.-C., GEIL L., ORCUTT M.L., STACKHOUSE T., LIPAN J., SLIFE L.,
RA BRAUCH H., DECKER J., NIEHANS G., HUGHSON M.D., MOCH H., STORKEL S.,
RA LERMAN M.I., LINEHAN W.M., ZBAR B.;
RL NAT. GENET. 16:68-73(1997).
CC CC FUNCTION: RECEPTOR FOR HEPATOCYTE GROWTH FACTOR. HAS A TYROSINE-
CC CC PROTEIN KINASE ACTIVITY.
CC CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC CC PROTEIN TYROSINE PHOSPHATE.
CC CC -!- SUBUNIT: HETERODIMER FORMED OF AN ALPHA CHAIN (50 KD) AND A BETA
CC CC CHAIN (145 KD) WHICH ARE DISULFIDE LINKED.
CC CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC CC -!- DISEASE: ACTIVATION OF MET AFTER REARRANGEMENT WITH THE TPR
CC CC GENE PRODUCES AN ONCOGENIC PROTEIN.
CC CC -!- DISEASE: DEFECTS IN MET ARE THE CAUSE OF HEREDITARY PAPILLARY
CC CC RENAL CARCINOMA (HPRC). HPRC IS A FORM OF INHERITED KIDNEY CANCER
CC CC CHARACTERIZED BY A PREDISPOSITION TO DEVELOP MULTIPLE, BILATERAL
CC CC PAPILLARY RENAL TUMORS. THE PATTERN OF INHERITANCE IS CONSISTENT
CC CC WITH AUTOSOMAL DOMINANT TRANSMISSION WITH REDUCED PENETRANCE.
CC CC -!- SIMILARITY: BELONGS TO THE MET TYROSINE KINASE FAMILY OF RECEPTOR.
CC EMBL; M35074; G386868; -.

DR EMBL; X54559; -; NOT_ANNOTATED_CDS.
DR EMBL; J02958; G307196; -.
DR EMBL; AC002080; G2078456; -.
DR PIR; A40175; TVHUME.
DR MIM; 164860; -.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR TRANSFERASE; TYROSINE-PROTEIN KINASE; PROTO-ONCOGENE; ATP-BINDING;
KW RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN; PHOSPHORYLATION; SIGNAL;
KW CHROMOSOMAL TRANSLOCATION; DISEASE MUTATION; POLYMORPHISM.
FT SIGNAL 1 24
FT CHAIN 25 1390 HEPATOCYTE GROWTH FACTOR RECEPTOR.
FT DOMAIN 25 932 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 933 955 POTENTIAL.
FT DOMAIN 956 1390 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1078 1345 PROTEIN KINASE.
FT NP_BIND 1084 1092 ATP (BY SIMILARITY).
FT BINDING 1110 1110 ATP (BY SIMILARITY).
FT ACT_SITE 1204 1204 BY SIMILARITY.
FT SITE 307 308 CLEAVAGE (POTENTIAL).
FT SITE 1009 1010 BREAKPOINT FOR TRANSLOCATION TO FORM
FT MOD_RES 1235 1235 TPR-MET ONCOGENE.
FT CARBOHYD 45 45 PHOSPHORYLATION (AUTO-).
FT CARBOHYD 106 106 POTENTIAL.
FT CARBOHYD 149 149 POTENTIAL.
FT CARBOHYD 202 202 POTENTIAL.
FT CARBOHYD 399 399 POTENTIAL.
FT CARBOHYD 405 405 POTENTIAL.
FT CARBOHYD 607 607 POTENTIAL.
FT CARBOHYD 635 635 POTENTIAL.
FT CARBOHYD 785 785 POTENTIAL.
FT CARBOHYD 879 879 POTENTIAL.
FT CARBOHYD 930 930 POTENTIAL.
FT VARIANT 320 320 A -> V.
FT VARIANT 1131 1131 M -> L (IN HPRC; GERMLINE MUTATION).
FT VARIANT 1188 1188 V -> L (IN HPRC; GERMLINE MUTATION).
FT VARIANT 1195 1195 L -> V (IN HPRC; SOMATIC MUTATION).
FT VARIANT 1220 1220 V -> I (IN HPRC; SOMATIC MUTATION).
FT VARIANT 1228 1228 D -> N (IN HPRC; GERMLINE MUTATION).
FT VARIANT 1228 1228 D -> H (IN HPRC; SOMATIC MUTATION).
FT VARIANT 1230 1230 Y -> C (IN HPRC; GERMLINE MUTATION).
FT VARIANT 1230 1230 Y -> H (IN HPRC; SOMATIC MUTATION).
FT VARIANT 1250 1250 M -> T (IN HPRC; SOMATIC MUTATION).
FT CONFLICT 755 755 S -> STWKEPLNIVSFLFCFAS (IN REF. 2).
FT CONFLICT 1191 1191 G -> A (IN REF. 2).
SQ SEQUENCE 1390 AA; 155526 MW; 650992C2 CRC32;
Query Match 70.5%; Score 43; DB 1; Length 1390;
Best Local Similarity 60.0%; Pred. No. 7.43e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 348 QSKPDSAEPM 357
QY 1 QENPDSSEPV 10
I :|||:|:
RESULT 12
ID LMA4_HUMAN STANDARD; PRT; 1816 AA.
AC Q16363; Q15335; Q14735;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE LAMININ ALPHA-4 CHAIN PRECURSOR.
GN LAMA4.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FETAL LUNG.
RX MEDLINE; 95300971.

RA IIVANAINEN A., SAINIO K., SARIOLA H., TRYGGVASON K.;
 RN FEBS LETT. 365:183-188(1995).
 [2]
 RN SEQUENCE FROM N.A.
 RC TISSUE=HEART;
 RA RICHARDS A.J., AL-IMARA L., POPE F.M.;
 RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 [3]
 RN SEQUENCE OF 236-1816 FROM N.A.
 RP
 RC TISSUE=HEART;
 RX MEDLINE; 95048381.
 RA RICHARDS A.J., AL-IMARA L., CARTER N.P., LLOYD J.C., LEVERSHA M.A.,
 RA POPE F.M.;
 RL GENOMICS 22:237-239(1994).
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR. LAMININ
 IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
 CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
 WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
 DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
 TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
 COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
 MEMBRANES (MAJOR COMPONENT).
 CC -1- TISSUE SPECIFICITY: IN ADULT, STRONG EXPRESSION IN HEART, LUNG,
 OVARY SMALL AND LARGE INTESTINES, PLACENTA, LIVER; WEAK OR NO
 EXPRESSION IN SKELETAL MUSCLE, KIDNEY, PANCREAS, TESTIS, PROSTATE,
 BRAIN. HIGH EXPRESSION IN FETAL LUNG AND KIDNEY. EXPRESSION IN
 FETAL AND NEWBORN TISSUES IS OBSERVED IN CERTAIN MESENCHYMAL CELLS
 IN TISSUES SUCH AS SMOOTH MUSCLE AND DERMIS.
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAIN G IS GLOBULAR.
 CC -1- SIMILARITY: CONTAINS 3.5 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
 CC -1- CAUTION: GENE LAMA4 WAS FORMERLY CALLED LAMA3.
 DR EMBL; S78569; G1042082;
 DR EMBL; X91171; E198045;
 DR EMBL; X76939; G509806;
 DR MIM; 600133;
 DR PROSITE; PS00222; EGF_1; 1.
 DR PROSITE; PS01248; LAMININ TYPE EGF; 3.
 KW GLYCOPROTEIN: BASEMENT MEMBRANE; EXTRACELLULAR MATRIX; COILED COIL;
 KW LAMININ EGF-LIKE DOMAIN; CELL ADHESION; REPEAT; SIGNAL.
 FT SIGNAL 1 24
 FT CHAIN 25 1816
 FT DOMAIN 82 255
 FT
 FT DOMAIN 82 131
 FT DOMAIN 132 186
 FT DOMAIN 187 240
 FT DOMAIN 241 255
 FT DOMAIN 256 851
 FT DOMAIN 852 1816
 FT DOMAIN 852 1027
 FT DOMAIN 1028 1219
 FT DOMAIN 1220 1449
 FT DOMAIN 1450 1632
 FT DOMAIN 1633 1816
 FT DOMAIN 313 396
 FT DOMAIN 466 521
 FT DOMAIN 574 607
 FT DOMAIN 655 717
 FT DOMAIN 770 799
 FT SITE 717 719
 FT SITE 82 91
 FT DISULFID 84 98
 FT DISULFID 101 110
 FT DISULFID 113 129
 FT DISULFID 132 146
 FT DISULFID 134 155
 FT DISULFID 157 166
 FT DISULFID 169 184

FT DISULFID 187 202
 FT DISULFID 189 209
 FT DISULFID 212 221
 FT DISULFID 224 238
 FT DISULFID 266 266
 FT DISULFID 269 269
 FT CARBOHYD 104 104
 FT CARBOHYD 215 215
 FT CARBOHYD 308 308
 FT CARBOHYD 458 458
 FT CARBOHYD 524 524
 FT CARBOHYD 550 550
 FT CARBOHYD 571 571
 FT CARBOHYD 574 574
 FT CARBOHYD 631 631
 FT CARBOHYD 639 639
 FT CARBOHYD 735 735
 FT CARBOHYD 751 751
 FT CARBOHYD 754 754
 FT CARBOHYD 780 780
 FT CARBOHYD 803 803
 FT CARBOHYD 1086 1086
 FT CARBOHYD 1281 1281
 FT CARBOHYD 1359 1359
 FT CARBOHYD 1411 1411
 FT CONFLICT 143 143
 FT CONFLICT 178 178
 FT CONFLICT 491 491
 FT CONFLICT 1057 1057
 SQ SEQUENCE 1816 AA; 201964 MW; C31FEBC CRC32;
 Query Match 70.5%; Score 43; DB 1; Length 1816;
 Best Local Similarity 55.6%; Pred. No. 7.43e+00;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 45 QDPPTSEP 53
 Qy 1 QENPDSSEP 9

RESULT 13
 ID SKN7 YEAST STANDARD; PRT; 622 AA.
 AC P38889; P39747;
 DT 01-FEB-1995 (REL. 31, CREATED)
 DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE PUTATIVE TRANSCRIPTION FACTOR SKN7 (POS9 PROTEIN).
 GN SKN7 OR POS9 OR YHR206W.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE; 94042854.
 RA BROWN J.L., NORTH S., BUSSEY H.;
 RL J. BACTERIOL. 175:6908-6915(1993).
 RN [2]
 RN SEQUENCE FROM N.A.
 RA KREMS B., CHARIZANIS C., ENTIAN K.-D.;
 RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RX MEDLINE; 94378003.
 RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DOVER J.;
 RA DU Z., FAVELLO A., FULTON L., GATTUNG S., GEISEL C., KIRSTEN J.,
 RA KUCABA T., HILLIER L., JIER M., JOHNSTON L., LANGSTON Y.,
 RA LATREILLE P., LOUIS E.J., MACRI C., MARDIS E., MENEZES S., MOUSER L.,
 RA NHAN M., RIFKIN L., RILES L., ST PETER H., TREVASKIS E., VAUGHAN K.,
 RA VIGNATI D., WILCOX L., WOHLDMAN P., WATERSTON R., WILSON R.,
 RA VAUDIN M.;
 RL SCIENCE 265:2077-2082(1994).
 RN [4]
 RN FUNCTION, AND MUTAGENESIS.

```

RX MEDLINE; 95045411.
RA BROWN J.L., BUSSEY H., STEWART R.C.;
RL EMBO J. 13:5186-5194(1994).
CC -!- FUNCTION: INVOLVED IN OXIDATIVE STRESS, TRANSCRIPTION FACTOR THAT
CC MAY FUNCTION IN A TWO-COMPONENT SIGNAL/TRANSDUCTION PATHWAY THAT
CC ACTS IN PARALLEL WITH THE PKC1 CASCADE TO REGULATE GROWTH AT THE
CC CELL SURFACE.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE HSF FAMILY.
CC -!- SIMILARITY: TO PROKARYOTE SENSORY TRANSDUCTION PROTEINS.
DR EMBL; U00485; G414419;
DR EMBL; X83031; G600028;
DR EMBL; U00029; G458922;
DR PIR; A49344; A49344.
DR PIR; S48987; S48987.
DR SGD; L0001908; SKN7.
DR PROSITE; PS00434; HSF_DOMAIN; 1.
KW TRANSCRIPTION REGULATION; SENSORY TRANSDUCTION; NUCLEAR PROTEIN;
FT DNA_BIND 86 190 BY SIMILARITY.
FT DNA_BIND 427 427 PHOSPHORYLATION (PROBABLE).
FT MOD_RES 427 427 D->N: DIMINISHED ACTIVITY.
FT MUTAGEN 427 427 D->E: AUGMENTED ACTIVITY.
FT MUTAGEN 427 427
SQ SEQUENCE 622 AA; 69202 MW; 4E506931 CRC32;

Query Match 68.9%; Score 42; DB 1; Length 622;
Best Local Similarity 50.0%; Pred. No. 1.26e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 536 QDNPSTTPV 545
QY 1 QENPDSSEPV 10

RESULT 14
ID TSGA_RAT STANDARD; PRT; 1214 AA.
AC Q63679;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TESTIS SPECIFIC PROTEIN A (ZINC FINGER PROTEIN TSGA).
GN TSGA.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE DAWLEY; TISSUE-TESTIS;
RX MEDLINE; 92172411
RA HOOG C., SCHALLING M., BRUNDELL E., DANEHOLT B.;
RL MOL. REPROD. DEV. 30:173-181(1991).
CC -!- FUNCTION: MEIOTIC OR POSTMEIOTIC FUNCTION.
CC -!- TISSUE SPECIFICITY: TESTIS-SPECIFIC; EXPRESSED ONLY IN MALE GERM
CC CELLS.
CC -!- DEVELOPMENTAL STAGE: REACHES A MAXIMUM DURING THE MEIOTIC AND THE
CC POSTMEIOTIC STAGES OF GERM CELL DEVELOPMENT.
DR EMBL; X59993; G57504;
KW ZINC-FINGER; METAL-BINDING.
FT ZN_FING 546 571 C6-TYPE.
FT ZN_FING 546 571
SQ SEQUENCE 1214 AA; 135403 MW; 05B39332 CRC32;

Query Match 68.9%; Score 42; DB 1; Length 1214;
Best Local Similarity 50.0%; Pred. No. 1.26e+01;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 1084 QENPADHDPI 1093
QY 1 QENPDSSEPV 10

RESULT 15
ID SCS2_YEAST STANDARD; PRT; 244 AA.
AC P40075;

```

```

DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE SCS2 PROTEIN
GN SCS2 OR YER120W.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RA NIKAWA J.-I., MURAKAMI A., ESUMI E., HOSAKA K.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA DIETRICH F.S., MULLIGAN J.T., HENNESSEY K.M., ALLEN E., ARAUJO R.,
RA AVILES E., BERNO A., BRENNAN T., CARPENTER J., CHEN E., CHERRY J.M.,
RA CHONG E., DUNCAN M., GUZMAN E., HARTZELL G., HUNICKE-SMITH S.,
RA HYMAN R., KAYSER A., KOMP C., LASHKARI D., LEW H., LIN D.,
RA MOSEDALE D., NAKAHARA K., NAMATH A., NORGREN R., OEFNER P., OH C.,
RA PETEL F.X., ROBERTS D., SEHL P., SCHRAMM S., SHOGREN T., SMITH V.,
RA TAYLOR P., WEI Y., YELTON M., BOISTEIN D., DAVIS R.W.;
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: CAN SUPPRESS AN INOSITOL AUXOTROPHIC MUTANT AND A
CC CHOLINE SENSITIVE MUTANT.
DR EMBL; D44493; G624933;
DR EMBL; U18916; G603359;
DR SGD; L0002629; SCS2.
SQ SEQUENCE 244 AA; 26925 MW; DA1802FD CRC32;

```

```

Query Match 67.2%; Score 41; DB 1; Length 244;
Best Local Similarity 60.0%; Pred. No. 2.11e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

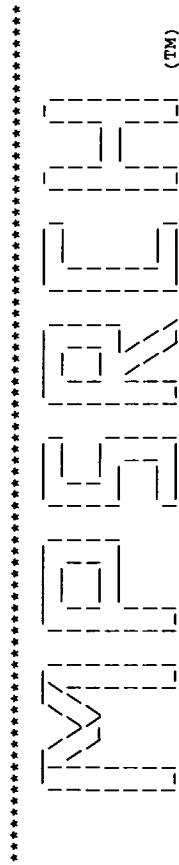
Db 140 QENKETVEPV 149
QY 1 QENPDSSEPV 10

```

```

Search completed: Thu May 13 15:45:37 1999
Job time : 8 secs.

```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 13 15:44:52 1999; Maspar time 2.92 Seconds
Tabular output not generated. 128.268 Million cell updates/sec

Title: >US-09-040-485-9
Description: (1-10) from US09040485.pep
Perfect Score: 61
Sequence: 1 QENPDSEPV 10

Scoring table: PAM 150
Gap 15

Searched: 116738 seqs, 37463448 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir58
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 20.371; Variance 24.321; scale 0.838

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description	Pred. No.
1	61	100.0	754	1 BABOH	peptide-aspartate bet	1.20e-03
2	61	100.0	757	2 I38423	aspartyl beta-hydroxy	1.20e-03
3	48	78.7	1379	2 S01254	hepatocyte growth fac	1.55e-00
4	46	75.4	440	2 B71293	hypothetical protein	4.28e-00
5	43	70.5	184	2 A61628	early gland protein e	1.87e-01
6	43	70.5	278	1 TPDUFW	troponin T, slow skel	1.87e-01
7	43	70.5	395	2 JQ0430	hypothetical 44.4K pr	1.87e-01
8	43	70.5	1218	2 S38182	probable transport pr	1.87e-01
9	43	70.5	1390	1 TVHWE	hepatocyte growth fac	1.87e-01
10	43	70.5	1816	1 S68960	laminin alpha-4 chain	1.87e-01
11	42	68.9	167	2 E71016	hypothetical protein	3.01e-01
12	42	68.9	607	2 S27776	80K protein (allele C	3.01e-01
13	42	68.9	622	1 A49344	cell wall assembly re	3.01e-01
14	42	68.9	634	2 JC4248	calcium binding FW29	3.01e-01
15	42	68.9	1214	2 S28499	probable finger prote	3.01e-01
16	41	67.2	200	2 H70409	conserved hypothetical	4.81e-01
17	41	67.2	244	2 S50623	SCS2 protein - yeast	4.81e-01
18	41	67.2	311	2 S06619	syndecan-1 precursor	4.81e-01
19	41	67.2	313	2 A42833	syndecan core protein	4.81e-01
20	41	67.2	338	2 P09437	hypothetical protein	4.81e-01
21	41	67.2	539	2 B56447	CTP synthetase homolo	4.81e-01
22	41	67.2	539	2 H71545	probable ctp syntheta	4.81e-01
23	41	67.2	727	2 A56879	diacylglycerol kinase	4.81e-01

24 41 67.2 938 1 Q0BE24 nuclear antigen EBNA- 4.81e+01
25 41 67.2 946 2 S27921 nuclear antigen EBNA- 4.81e+01
26 41 67.2 1857 2 S01787 fatty-acid synthase (4.81e+01
27 40 65.6 151 2 I38243 zona binding protein 7.61e+01
28 40 65.6 204 2 B71433 hypothetical protein 7.61e+01
29 40 65.6 215 2 JG1133 alpha-s1-casein precu 7.61e+01
30 40 65.6 319 2 S44642 F37A4.5 protein - Cae 7.61e+01
31 40 65.6 355 2 A44851 3-isopropylmalate deh 7.61e+01
32 40 65.6 393 2 S60465 dom-3 protein - Caeno 7.61e+01
33 40 65.6 415 2 S39195 levansucrase - Erwin 7.61e+01
34 40 65.6 429 2 S72829 trehalase-6-phosphate 7.61e+01
35 40 65.6 459 2 B44498 radial spoke protein 7.61e+01
36 40 65.6 472 2 I49590 CD36 antigen - mouse 7.61e+01
37 40 65.6 472 1 A47402 fatty acid binding/tr 7.61e+01
38 40 65.6 484 1 K1ECXY xylulokinase (EC 2.7 7.61e+01
39 40 65.6 586 2 B70503 probable pyrg protein 7.61e+01
40 40 65.6 685 2 S64158 hypothetical protein 7.61e+01
41 40 65.6 702 2 E69498 hypothetical protein 7.61e+01
42 40 65.6 759 2 G59258 ATP-dependent RNA hel 7.61e+01
43 40 65.6 774 2 JG6095 hepatocyte nuclear fa 7.61e+01
44 40 65.6 1350 2 G36793 hypothetical protein 7.61e+01
45 40 65.6 1375 2 JG5148 hepatocyte growth fac 7.61e+01

ALIGNMENTS

RESULT 1
ENTRY BABOH #type complete
TITLE peptide-aspartate beta-dioxygenase (EC 1.14.11.16) - bovine
ALTERNATE_NAMES aspartyl (asparaginyl) beta-hydroxylase
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 31-Dec-1993 #sequence_revision 10-Feb-1995 #text_change
29-May-1998
ACCESSIONS A42969; A39470; B39470; C39470; S27948
REFERENCE A42969
#authors ~~Stern, A.M.; VanDusen, W.J.; Dietz, R.E.; Kohler, N.E.; Dixon, R.A.; Elliston, K.O.; Stern, A.M.; Friedland, P.A.~~
#journal J. Biol. Chem. (1992) 267:14322-14327
#title CDNA cloning and expression of bovine aspartyl (asparaginyl) beta-hydroxylase.
#cross-references MIM:92332546
#accession A42969
#molecule_type mRNA
#residues 1-754 #label J1A
#cross-references EMBL:M91213; NID:g162693; PID:g162694
#experimental_source brain
#note sequence extracted from NCBI backbone (NCBIP:108534)
REFERENCE A39470
#authors Wang, Q.; VanDusen, W.J.; Petroski, C.J.; Garsky, V.M.; Stern, A.M.; Friedland, P.A.
#journal J. Biol. Chem. (1991) 266:14004-14010
#title Bovine liver aspartyl beta-hydroxylase. Purification and characterization.
#cross-references MIM:91310689
#accession A39470
#molecule_type protein
#residues 289-328 #label WAN
#accession B39470
#molecule_type protein
#residues 615,'X',617-630,'XX',633-634,'X',636,'XX',639-641
#accession C39470
#molecule_type protein
#residues 311-347,'X',349,'X',351-373,'X',375-379,'X',381-382
#label WA3

COMMENT This enzyme uses ferrous iron as a cofactor, and while beta-hydroxylating the peptidyl-aspartate substrate converts alpha-ketoglutarate to succinate and releases carbon dioxide.
COMMENT Aspartic acid and asparagine residues in the EGF homology domain of certain plasma proteins serve as the peptidyl-aspartate substrate.
CLASSIFICATION #superfamily peptide-aspartate beta-dioxygenase; tetratricopeptide repeat homology

```

KEYWORDS
FEATURE
2-56
57-78
289-754
311-754
337-370
371-404
13,96,466,702

#domain intracellular #status predicted #label INC\
#domain transmembrane #status predicted #label TRM\
#product peptide-aspartate beta-dioxygenase, 56K form
#status predicted #label 56K\
#product peptide-aspartate beta-dioxygenase, 52K form
#status predicted #label 52K\
#domain tetratricopeptide repeat homology #label TT1\
#domain tetratricopeptide repeat homology #label TT2\
#binding_site carboxylate (Asn) (covalent) #status
predicted
#length 754 #molecular-weight 84998 #checksum 9667

SUMMARY
Query Match 100.0%; Score 61; DB 1; Length 754;
Best Local Similarity 100.0%; Pred. No. 1.20e-03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 248 QENPDSSEPV 257
|||||
QY 1 QENPDSSEPV 10

RESULT 2
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:95121937
#accession
#status
#molecule_type mRNA
#residues 1-757 #label RES
#cross-references EMBL:U03109; NID:q458032
CLASSIFICATION
#superfamily peptide-aspartate beta-dioxygenase;
tetratricopeptide repeat homology
FEATURE
54-75
SUMMARY
#domain transmembrane #status predicted #label TRM
#length 757 #molecular-weight 85498 #checksum 2143

Query Match 100.0%; Score 61; DB 2; Length 757;
Best Local Similarity 100.0%; Pred. No. 1.20e-03;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db      233 QENPDSSEPV 242
        |||||
QY      1 QENPDSSEPV 10

RESULT 3
ENTRY
TITLE   S01254 #type complete
CONTAINS
  hepatocyte growth factor receptor precursor - mouse
  protein-tyrosine kinase (EC 2.7.1.112) met
  #formal_name Mus musculus #common name house mouse
  30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change
  08-Sep-1997
ACCESSIONS
REFERENCE
  S01254; JH0115; A45453
  S01254
  Chan, A.M.L.; King, H.W.S.; Deakin, E.A.; Tempest, P.R.;
  Hilkenes, J.; Kroezen, V.; Edwards, D.R.; Wills, A.J.;
  Brookes, P.; Cooper, C.S.
  #journal      Oncogene (1988) 2:593-599
  #title        Characterization of the mouse met proto-oncogene.
  #cross-references MUID:88262253
  #accession    S01254
  #molecule_type mRNA

```

```

##residues      1-1379  ##label CHA
##cross-references EMBL:Y00671; NID:g53058; PID:g53059
REFERENCE      JH0112
#authors      Wilks, A.F.; Kurban, R.R.; Hovens, C.M.; Ralph, S.J.
#journal      Gene (1989) 85:67-74
#title        The application of the polymerase chain reaction to cloning
               members of the protein tyrosine kinase family.
#cross-references MUID:90152381
#accession    JH0115
##molecule_type mRNA
##residues     'I',1200-1254,'R',1256-1260,'T',1262-1268  ##label WIL
##experimental_source hemopoietic cell
##note         the authors translated the codon ACG for residue 1261 as
               Lys
REFERENCE      A45453
#authors      Weidner, K.M.; Sachs, M.; Birschmeier, W.
#journal      J. Cell Biol. (1993) 121:145-154
#title        The Met receptor tyrosine kinase transduces motility,
               proliferation, and morphogenic signals of scatter
               factor/hepatocyte growth factor in epithelial cells.
#accession    A45453
##status      preliminary; not compared with conceptual translation
##molecule_type mRNA
##residues     924-935  ##label WEI

GENETICS
#gene          met
#superfamily  hepatocyte growth factor receptor; protein
               kinase homology
#ATP:          autophosphorylation; glycoprotein; phosphoprotein;
               phosphotransferase; proto-oncogene; receptor; transmembrane
               protein; tyrosine-specific protein kinase

FEATURE
1-24          #domain signal sequence #status predicted #label SIG\
25-929        #domain extracellular #status predicted #label EXT\
25-302        #product hepatocyte growth factor receptor alpha chain
               #status predicted #label ACH\
308-1379      #product hepatocyte growth factor receptor beta chain
               #status predicted #label BCH\
930-954       #domain transmembrane #status predicted #label TM\
955-1379      #domain intracellular #status predicted #label INT\
1074-1342     #domain protein kinase homology #label KIN\
1082-1090     #region protein kinase ATP-binding motif\
1108          #active_site Lys #status predicted\
1233          #binding_site phosphate (Tyr) (covalent) (by
               autophosphorylation) #status predicted
               #length 1379 #molecular-weight 153547 #checksum 6033

SUMMARY
Query Match      78.7%; Score 48; DB 2; Length 1379;
Best Local Similarity 70.0%; Pred. No. 1.55e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

Db	347 QSKPDSAEPPV	356
	I : :	
QY	1 QENPDSEPPV	10
RESULT	4	
ENTRY	B71293	#type complete
TITLE	hypothetical protein TP0693 - syphilis spirochete	
ORGANISM	#formal_name Treponema pallidum subsp. pallidum #common_name syphilis spirochete	
DATE	24-Jul-1998	#sequence_revision 24-Jul-1998 #text_change 07-Aug-1998
ACCESSIONS	B71293	
REFERENCE	A71250	
#authors	Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwinn, M.; Hickey, E.K.; Clayton, R.; Ketchum, K.A.; Sodergren, E.; Hardham, J.M.; McLeod, M.P.; Salzberg, S.; Peterson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDonald, L.; Attalach, P.; Bowman, C.; Cotton, M.D.; Fujii, C.; Garland, S.; Hatch, B.; Horst, K.; Roberts, K.; Watthey, L.; Weidman, J.; Smith, H.O.; Venter, J.C.	

```

#journal Science (1998) 281:375-388
#title Complete genome sequence of Treponema pallidum, the syphilis
#accession B71293
#status preliminary; nucleic acid sequence not shown;
      translation not shown
#molecule_type DNA
##residues 1-440 ##label COL
##cross-references GB:AE001243; GB:AE000520; NID:g3322990; PID:g3323000
##experimental_source strain Nichols
GENETICS
#gene TP0693
SUMMARY
#length 440 #molecular-weight 47672 #checksum 4753
Query Match 75.4%; Score 46; DB 2; Length 440;
Best Local Similarity 70.0%; Pred. No. 4.28e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 244 QERPSSPEPV 253
   ||| |||
QY 1 QENPDSPEPV 10

RESULT 5
ENTRY #type complete
TITLE early gland protein egp-1 precursor - fruit fly (Drosophila
      virilis)
ORGANISM #formal_name Drosophila virilis
DATE 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change
REFERENCE A61628; S57581
#authors Thueroff, E.; Stoeven, S.; Kress, H.
#journal Mech. Dev. (1992) 37:81-93
#title Drosophila salivary glands exhibit a regional reprogramming
      of gene expression during the third larval instar.
#accession A61628
#status preliminary; not compared with conceptual translation
##molecule_type mRNA
##residues 1-184 ##label THU
REFERENCE S57581
#authors Stoeven, S.
#submission submitted to the EMBL Data Library, June 1995
#accession S57581
#status preliminary
##molecule_type DNA
##residues 1-184 ##label STO
##cross-references EMBL:249942; NID:g887425; PID:g887426
GENETICS
#gene FlyBase:Dvir/Egpl
#keywords extracellular protein; salivary gland
SUMMARY #length 184 #molecular-weight 20567 #checksum 4781
Query Match 70.5%; Score 43; DB 2; Length 184;
Best Local Similarity 50.0%; Pred. No. 1.87e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 77 EESPEDEPV 86
   ||| |||
QY 1 QENPDSPEPV 10

RESULT 6
ENTRY #type complete
TITLE troponin T, slow skeletal muscle - human
ALTERNATE_NAMES troponin T1
ORGANISM #formal_name Homo sapiens #common_name man
DATE 20-Jun-1989 #sequence_revision 17-May-1996 #text_change
ACCESSIONS A29783; B29783; JC2126; A57979
REFERENCE A29783
#authors Gahlmann, R.; Troutt, A.B.; Wade, R.P.; Gunning, P.; Kedes,

```

```

L.
#journal J. Biol. Chem. (1987) 262:16122-16126
#title Alternative splicing generates variants in important
      functional domains of human slow skeletal troponin T.
#cross-references MUID:88058976
#accession A29783
#molecule_type mRNA
##residues 1-19,'D',21-278 ##label GAH
##cross-references GB:M19309; NID:g339780; PID:g339781; GB:J03476
##experimental_source clone H22h
#accession B29783
#molecule_type mRNA
##residues 1-24,36-204,221-278 ##label GA2
##cross-references GB:M19308; NID:g339782; PID:g339783; GB:J03476
#experimental_source clone M1
REFERENCE JC2126
#authors Samson, F.; Mesnard, L.; Mihovilovic, M.; Potter, T.G.;
      Mercadier, J.J.; Roses, A.D.; Gilbert, J.R.
#journal Biochem. Biophys. Res. Commun. (1994) 199:841-847
#title A new human slow skeletal troponin T (TnTs) mRNA isoform
      derived from alternative splicing of a single gene.
#accession JC2126
#molecule_type mRNA
##residues 1-204,221-278 ##label SAM
##cross-references GB:S69208; NID:g546020; PID:g546021
#accession A57979
#molecule_type mRNA
##residues 1-24,36-204,221-278 ##label SA2
##cross-references GB:S69209; NID:g546022; PID:g546023
GENETICS
#gene GDB:TNNT1
##cross-references GDB:125310; OMIM:191041
#map_position 19q13.4-19q13.4
#complex troponin is a heterotrimer with one molecule each of troponin
      C (calcium binding component), troponin I (inhibitory
      component), and troponin T (tropomyosin-binding component)
FUNCTION
#description binds the troponin complex to tropomyosin; with tropomyosin
      mediates contraction of vertebrate striated muscle in
      response to calcium
#pathway muscle contraction
#classification #superfamily troponin T
#keywords acetylated amino end; actin binding; alternative splicing;
      muscle contraction; phosphoprotein; skeletal muscle; thin
      filaments
FEATURE
2-278 #product troponin T, slow skeletal muscle splice form 1
      #status predicted #label MAR1\
2-204,221-278 #product troponin T, slow skeletal muscle splice form 2
      #status predicted #label MAR2\
2-24,36-204,
221-278 #product troponin T, slow skeletal muscle splice form 3
      #status predicted #label MAR3\
2 #modified_site acetylated amino end (Ser) (in mature
      form) #status predicted\
2 #binding_site phosphate (Ser) (covalent) (by troponin T
      kinase) #status predicted\
177 #binding_site phosphate (Thr) (covalent) (by
      calmodulin-dependent kinase II) #status predicted
SUMMARY #length 278 #molecular-weight 32948 #checksum 4551
Query Match 70.5%; Score 43; DB 1; Length 278;
Best Local Similarity 50.0%; Pred. No. 1.87e-01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 23 EEAPPEPEPV 32
   ||| |||
QY 1 QENPDSPEPV 10

RESULT 7
ENTRY #type complete
TITLE hypothetical 44.4K protein - Streptomyces fradiae transposon

```

```

Tn4556
#formal_name Streptomyces fradiae
#sequence_revision 07-Sep-1990 #text_change
18-Jun-1993
ACCESSIONS
REFERENCE JQ0430
#authors Sliemantiak, D.R.; Slightom, J.L.; Chung, S.T.
#journal Gene (1990) 86:1-9
#title Nucleotide sequence of Streptomyces fradiae transposable
#cross-references MUID:90185236
#accession JQ0430
#molecule_type DNA
#residues 1-395 #label SIE
SUMMARY #length 395 #molecular-weight 44379 #checksum 7125
Query Match 70.5%; Score 43; DB 2: Length 395;
Best Local Similarity 60.0%; Pred. No. 1.87e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 58 QBRADGAEVP 67
|| :||:||||
Qy 1 QENPDSEVP 10

RESULT 8
ENTRY S38182 #type complete
TITLE probable transport protein YKR103w - yeast (Saccharomyces
cerevisiae)
ALTERNATE_NAMES multidrug resistance protein homolog YKR103w
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 03-May-1994 #sequence_revision 03-May-1994 #text_change
24-Jul-1998
ACCESSIONS S38182
REFERENCE S38175
#authors Gaillon, L.; Dujon, B.
#submission submitted to the Protein Sequence Database, March 1994.
#accession S38182
#molecule_type DNA
#residues 1-1218 #label GAI
#cross-references EMBL:228328; NID:9486610; PID:9486611; MIPS:YKR103w
#experimental_source strain S288C
GENETICS
#map_position 11R
CLASSIFICATION #superfamily unassigned ATP-binding cassette proteins;
ATP-binding cassette homology
ATP; P-loop; transmembrane protein
KEYWORDS
FEATURE 659-868 #domain ATP-binding cassette homology #label ABC\
686-693 #region nucleotide-binding motif A (P-loop)\
692 #binding_site ATP (Lys) #status Predicted
SUMMARY #length 1218 #molecular-weight 137995 #checksum 1891
Query Match 70.5%; Score 43; DB 2: Length 1218;
Best Local Similarity 77.8%; Pred. No. 1.87e+01;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 416 EENPDSEEA 424
:|||||:
Qy 1 QENPDSEEP 9

RESULT 9
ENTRY TVHUME #type complete
TITLE hepatocyte growth factor receptor precursor - human
CONTAINS protein-tyrosine kinase (EC 2.7.1.112) met
ORGANISM #formal_name Homo sapiens #common_name man
DATE 31-Mar-1991 #sequence_revision 30-Sep-1992 #text_change
20-Mar-1998
ACCESSIONS A40175; A28303; A93749; A93369; A53761; I57632; A30008;
B24569
REFERENCE A40175
#authors Giordano, S.

```

```

#submission submitted to the EMBL Data Library, November 1990
#accession A40175
#molecule_type mRNA
#residues 1-1390 #label GIO
#cross-references EMBL:X54559
REFERENCE A28303
#authors Park, M.; Dean, M.; Kaul, K.; Braun, M.J.; Gonda, M.A.; Vande
Woude, G.
#journal Proc. Natl. Acad. Sci. U.S.A. (1987) 84:6379-6383
#title Sequence of MET protooncogene cDNA has features
Characteristic of the tyrosine kinase family of
growth-factor receptors.
#cross-references MUID:87317655
#accession A28303
#molecule_type mRNA
#residues 1-755, 'TWMKEPLNIVSFLFCFAS', 756-1190, 'A', 1192-1390
#cross-references GB:J02958; NID:g187558; PID:g307196
REFERENCE A93749
#authors Chan, A.M.L.; King, H.W.S.; Tempest, P.R.; Deakin, E.A.;
Cooper, C.S.; Brookes, P.
#journal Oncogene (1987) 1:229-233
#title Primary structure of the met protein tyrosine kinase domain.
#cross-references MUID:88143699
#accession A93749
#molecule_type mRNA
#residues 'VNRETCQSLRLEKLNQKALTEKNKELETAQDNIAIQSO',
'FTRKKELEAEKRDILNRLSQSELYLT', 1010-1271, 'L',
1273-1390 #label CHA
#cross-references GB:U08818; NID:g487741; PID:g487742
#note this activated met oncogene is the product of gene
rearrangement
REFERENCE A93369
#authors Dean, M.; Park, M.; Le Beau, M.M.; Robins, T.S.; Diaz, M.O.;
Rowley, J.D.; Blair, D.G.; Vande Woude, G.F.
#journal Nature (1985) 318:385-388
#title The human met oncogene is related to the tyrosine kinase
oncogenes
#cross-references MUID:86065462
#accession A93369
#molecule_type DNA
#residues 1267-1390 #label DEA
#cross-references GB:M35074; NID:g187555; PID:g386868
REFERENCE A53761
#authors Gambartta, G.; Pistoi, S.; Giordano, S.; Comoglio, P.M.;
Santoro, C.
#journal J. Biol. Chem. (1994) 269:12852-12857
#title Structure and inducible regulation of the human MET promoter.
#accession A53761
#molecule_type mRNA
#residues 1-14 #label GAM
REFERENCE A40179
#authors Ferracini, R.; Longati, P.; Naldini, L.; Vigna, E.; Comoglio,
P.M.
#journal J. Biol. Chem. (1991) 266:19558-19564
#title Identification of the major autophosphorylation site of the
Met/hepatocyte growth factor receptor tyrosine kinase.
#cross-references MUID:92011756
#contents annotation; autophosphorylation site
REFERENCE I57632
#authors Dean, M.; Park, M.; Vande Woude, G.F.
#journal Mol. Cell. Biol. (1987) 7:921-924
#title Characterization of the rearranged tpr-met oncogene
breakpoint.
#cross-references MUID:87144265
#accession I57632
#status translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 963-1009 #label RES
#cross-references GB:M15325; NID:g187531; PID:g187532
COMMENT The receptor is a dimer of disulfide-bonded 50K alpha and 145K beta
chains that arise by cleavage of the precursor. Activity is
regulated by phosphorylation of serine and tyrosine residues.

```

```

GENETICS
#gene GDB:MET
#map_position 7q31-7q31
#cross-references GDB:120178; OMIM:164860
CLASSIFICATION #superfamily hepatocyte growth factor receptor; protein
#map_position 7q31-7q31 kinase homology
KEYWORDS ATP; autophosphorylation; glycoprotein; phosphoprotein;
phosphotransferase; proto-oncogene; receptor; transmembrane
protein; tyrosine-specific protein kinase
FEATURE
1-24 #domain signal sequence #status predicted #label SIG\
25-303 #product hepatocyte growth factor receptor alpha chain
#status predicted #label ALP\
308-1390 #product hepatocyte growth factor receptor beta chain
#status predicted #label BET\
933-955 #domain transmembrane #status predicted #label TMN\
1076-1344 #domain protein kinase homology #label KIN\
1084-1092 #region protein kinase ATP-binding motif\
45,106,149,202,399, #binding_site carbohydrate (Asn) (covalent) #status
405,635,785,930 predicted\
1110 #active_site lys #status experimental\
1235 #binding_site phosphate (Tyr) (covalent) (by
autophosphorylation) #status experimental
SUMMARY #length 1390 #molecular-weight 155526 #checksum 2959
Query Match 70.5%; Score 43; DB 1; Length 1390;
Best Local Similarity 60.0%; Pred. No. 1.87e+01;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 348 QSKPDSAEPM 357
QY 1 QENPDSSEP 10
RESULT 10
ENTRY S68960 #type complete
TITLE laminin alpha-4 chain precursor - human
ALTERNATE_NAMES laminin A4
ORGANISM Homo sapiens #common_name man
DATE 21-Aug-1998 #sequence_revision 21-Aug-1998 #text_change
ACCESSIONS S68960; S65926; S49149; S40150; I53516
REFERENCE S68960
#authors Richards, A.; Al-Imara, L.; Pope, F.M.
#journal Eur. J. Biochem. (1996) 238:813-821
#title The complete cDNA sequence of laminin alpha-4 and its
relationship to the other human laminin alpha chains.
#accession S68960
#molecule_type mRNA
#residues 1-1816 #label RIC
#cross-references EMBL:X91171; NID:g1212962; PID:e198045; PID:g1212963
#experimental_source tissue type heart
REFERENCE I53516
#authors Iivanainen, A.; Sainio, K.; Sariola, H.; Tryggvason, K.
#journal FEBS Lett. (1995) 365:183-188
#title Primary structure and expression of a novel human laminin
alpha-4 chain.
#cross-references MIM:95300971
#accession S65926
#molecule_type mRNA
#residues 1-142, 'P', 144-177, 'F', 179-490, 'Y', 492-1056, 'P', 1058-1816
#label IIV
#cross-references EMBL:S78569; NID:g1042081; PID:g1042082
REFERENCE S49149
#authors Richards, A.J.; Al-Imara, L.; Carter, N.; Leversha, M.;
Lloyd, J.C.; Pope, F.M.
#submission submitted to the EMBL Data Library, December 1993
#description Localisation of the gene (LAMA4) to chromosome 6q21 and
isolation of a partial cDNA encoding a variant laminin A
chain.
#accession S49149
#molecule_type mRNA

```

```

#residues 236-1816 #label RI2
#cross-references EMBL:X76939; NID:g509805; PID:g509806
REFERENCE S40150
#authors Richards, A.J.; Al-Imara, L.; Carter, N.; Lloyd, J.C.; Pope,
F.M.
#submission submitted to the EMBL Data Library, February 1993
#description Isolation of a partial cDNA encoding a protein homologous to
laminin A. Assignment of the gene to chromosome 6.
#accession S40150
#molecule_type mRNA
#residues 1403-1541, 'S', 1543-1816 #label RI3
#cross-references EMBL:X70904; NID:g437804; PID:g437805
GENETICS
#gene GDB:LAMA4; LAMA3
#cross-references GDB:203904; OMIM:600133
#map_position 6q21-6q21
COMPLEX Laminins are trimers of an alpha-type, a beta-type, and a
gamma-type laminin chain.
FUNCTION
#description interact with cells and with other basement membrane proteins
to promote differentiation, development, and cell migration
#superfamily laminin alpha-4 chain; laminin G repeat
homology; laminin-type EGF-like homology
KEYWORDS basement membrane; cell binding; coiled coil; extracellular
matrix; glycoprotein; heptad repeat; heterotrimer
FEATURE
1-24 #domain signal sequence #status predicted #label SIG\
25-1816 #product laminin alpha-4 chain #status predicted #label
MAT\
82-129 #domain laminin-type EGF-like homology #label LE1\
132-184 #domain laminin-type EGF-like homology #label LE2\
187-238 #domain laminin-type EGF-like homology #label LE3\
241-265 #domain laminin-type EGF-like homology #status atypical
#label LE4\
717-719 #region cell attachment motif (R-G-D)\
871-1003 #domain laminin G repeat homology #label LG1\
1068-1198 #domain laminin G repeat homology #label LG2\
1252-1367 #domain laminin G repeat homology #label LG3\
1488-1614 #domain laminin G repeat homology #label LG4\
1665-1789 #domain laminin G repeat homology #label LG5\
104,215,308,458,
524,550,571,574,
631,639,735,751,
754,780,803,1086,
1281,1359,1411 #binding_site carbohydrate (Asn) (covalent) #status
predicted\
266,269 #disulfide_bonds interchain #status predicted
SUMMARY #length 1816 #molecular-weight 201882 #checksum 8148
Query Match 70.5%; Score 43; DB 1; Length 1816;
Best Local Similarity 55.6%; Pred. No. 1.87e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 45 QDPETSEP 53
QY 1 QENPDSSEP 9
RESULT 11
ENTRY E71016 #type complete
TITLE hypothetical protein PH1427 - Pyrococcus horikoshii
ORGANISM #formal_name Pyrococcus horikoshii
DATE 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
14-Aug-1998
REFERENCE E71016
#accessions A71000.
#authors Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kishida, N.; Oguchi,
A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.

```

```
#journal      DNA Res. (1998) 5:55-76
#title       Complete sequence and gene organization of the genome of a
             hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
             OT3.
#accession   E71016
#status      preliminary; nucleic acid sequence not shown;
             translation not shown
#molecule_type DNA
##residues   1-167 ##label KAW
##cross-references GB:AP000006; NID:g3236133; PID:di031476; PID:g3257850
##experimental_source strain OT3
##note       this accession replaces an interim accession for a
             sequence replaced by GenBank
GENETICS
#gene        PH1427
#accession   S49986
#molecule_type DNA
#length      167 #molecular-weight 18831 #checksum 3158
SUMMARY
Query Match      68.9%; Score 42; DB 2; Length 167;
Best Local Similarity 85.7%; Pred. No. 3.01e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 148 PESSEPV 154
|:|||||
QY 4 PDSSEPV 10

RESULT 12
ENTRY   S27776      #type complete
TITLE   80K protein (allele C1B) - Babesia bovis
ORGANISM #formal_name Babesia bovis
DATE    17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change
09-Sep-1997
ACCESSIONS S27776
REFERENCE   S27776
#authors   Dalrymple, B.P.; Peters, J.M.
#submission submitted to the EMBL Data Library, May 1992
#description Sequence of cDNA clones of a babesia bovis gene isolated
             using sera from cattle vaccinated with a dextran sulphate
             antigen fraction.
#accession S27776
#molecule_type mRNA
##residues 1-607 ##label DAL
##cross-references EMBL:M93126; NID:g155862; PID:g155863
SUMMARY   #length 607 #molecular-weight 67129 #checksum 8353
Query Match      68.9%; Score 42; DB 2; Length 607;
Best Local Similarity 55.6%; Pred. No. 3.01e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 367 EEPEAEPV 375
|:|::|||
QY 2 ENPDSSEPV 10

RESULT 13
ENTRY   A49344      #type complete
TITLE   cell wall assembly regulatory protein SKN7 - yeast
             (Saccharomyces cerevisiae)
ALTERNATE_NAMES (Saccharomyces cerevisiae)
ORGANISM oxidative stress response regulator POS9; protein YHR206w
#formal_name Saccharomyces cerevisiae
DATE    07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change
17-Apr-1998
ACCESSIONS A49344; S48987; S49986; S68114
REFERENCE   A49344
#authors   Brown, J.L.; North, S.; Bussey, H.
#journal   J. Bacteriol. (1993) 175:6908-6915
#title     SKN7, a yeast multicopy suppressor of a mutation affecting
             cell wall beta-glucan assembly, encodes a product with
             domains homologous to prokaryotic two-component regulators
             and to heat shock transcription factors.
#accession A49344
#molecule_type DNA
##residues 1-622 ##label BRO
```

```
#cross-references GB:U00485; NID:g414418; PID:g414419
REFERENCE   S46671
#authors   Macri, C.
#submission submitted to the EMBL Data Library, February 1994
#description The sequence of S. cerevisiae cosmid 9177.
#accession S48987
#molecule_type DNA
##residues 1-622 ##label MAC
##cross-references EMBL:U00029; NID:g551322; PID:g458922; MIPS:YHR206w
REFERENCE   S49986
#authors   Krens, B.; Charizanis, C.; Entian, K.D.
#submission submitted to the EMBL Data Library, November 1994
#description A protein (Pos9) similar to prokaryotic response regulators
             is involved in oxidative stress in yeast.
#accession S49986
#molecule_type DNA
##residues 1-622 ##label KRE
##cross-references EMBL:X83031; NID:g600027; PID:g600028
REFERENCE   S68114
#authors   Krens, B.; Charizanis, C.; Entian, K.D.
#journal   Curr. Genet. (1996) 29:327-334
#title     The response regulator-like protein Pos9/Skn7 of
             Saccharomyces cerevisiae is involved in oxidative stress
             resistance.
#accession S68114
#status    nucleic acid sequence not shown; translation not shown
#molecule_type DNA
##residues 1-622 ##label KRW
##cross-references EMBL:X83031; NID:g600027; PID:g600028
##note     the nucleotide sequence was submitted to the EMBL Data
             Library, November 1994
GENETICS
#gene      SGD:SKN7; POS9
#accession SGD:S0001249; MIPS:YHR206w
#map_position 8R
CLASSIFICATION #superfamily cell wall assembly regulatory protein SKN7; HSF
               DNA-binding domain homology; response regulator homology
               DNA binding; leucine zipper; nucleus; phosphoprotein;
               transcription regulation
KEYWORDS
FEATURE
87-194      #domain HSF DNA-binding domain homology #label HSF\
379-488     #domain response regulator homology #label RRH\
382-410     #region leucine zipper\
555-576     #region glutamine-rich\
427         #binding-site phosphate (Asp) (covalent) #status
               predicted
SUMMARY   #length 622 #molecular-weight 69202 #checksum 256
Query Match      68.9%; Score 42; DB 1; Length 622;
Best Local Similarity 50.0%; Pred. No. 3.01e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 536 QDNPSTTPV 545
|:|::|||
QY 1 QENPDSSEPV 10

RESULT 14
ENTRY   JC4248      #type complete
TITLE   calcium binding PW29 protein - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE    12-Oct-1995 #sequence_revision 08-Feb-1996 #text_change
07-Nov-1997
ACCESSIONS JC4248
REFERENCE   JC4248
#authors   Yu, S.; Ozawa, M.; Naved, A.F.; Miyauchi, T.; Muramatsu, H.;
             Muramatsu, T.
#journal   Cell Struct. Funct. (1995) 20:263-268
#title     cDNA cloning and sequence analysis of a novel calcium binding
             protein with oligoproline motif.
#accession JC4248
#molecule_type mRNA
##residues 1-634 ##label YUS
```

##cross-references DBJ:D49429; NID:g699609; PID:d1009004; PID:gl304155
##experimental_source F9 embryonal carcinoma cells

COMMENT This protein is a cytoplasmic calcium binding protein which lacks EF-hand motif, and is present in embryonal carcinoma cells. It plays important roles in regulation of cellular activities. This protein is rich in hydrophilic amino acids.

KEYWORDS

calcium binding

FEATURE

528-547 #region glutamic acid/lysine-rich
SUMMARY #length 634 #molecular-weight 71893 #checksum 7243

Query Match

58.9%; Score 42; DB 2; Length 634;

Best Local Similarity 55.6%; Pred. No. 3.01e+01;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 276 DSPDSVDPV 284

QY 2 ENPDSSEPV 10

RESULT 15

ENTRY

S28499 #type complete

probable finger protein - rat

TITLE #formal_name Rattus norvegicus #common_name Norway rat

ORGANISM 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change

DATE 10-Sep-1997

ACCESSIONS

S28499

REFERENCE

#authors Hoog, C.; Schalling, M.; Grunder-Brundell, E.; Daneholt, B.

#submission submitted to the EMBL Data Library, June 1991

#description Analysis of a murine germ cell-specific transcript that

encodes a putative zinc finger protein.

#accession S28499

##molecule_type mRNA

##residues 1-1214 #label HOO

##cross-references EMBL:X59993; NID:g57503; PID:g57504

##experimental_source strain Sprague-Dawley

KEYWORDS DNA binding; zinc; zinc finger

SUMMARY #length 1214 #molecular-weight 135403 #checksum 4667

Query Match

68.9%; Score 42; DB 2; Length 1214;

Best Local Similarity 50.0%; Pred. No. 3.01e+01;

Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 1084 QENPADHDPI 1093

QY 1 QENPDSSEPV 10

Search completed: Thu May 13 15:45:09 1999

Job time : 17 secs.

MPSRCH_PP protein - protein database search, using Smith-Waterman algorithm

***** (TM) *****

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu May 13 15:45:58 1999; MasPar time 4.01 seconds
Tabular output not generated. 124.082 Million cell updates/sec

Title: >US-09-040-485-9
Description: (1-10) from US09040485.pep
Perfect Score: 61
Sequence: 1 QENPDSSEPV 10

Scoring table: PAM 150
Gap 15
Searched: 165420 seqs, 49795644 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrnbl6
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 20.152; Variance 22.779; scale 0.885

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	49	80.3	374	13	042322 D4B DOPAMINE RECEPTOR.	6.39e-01
2	46	75.4	344	5	062447 Y43F4B.1.	3.26e+00
3	44	72.1	590	6	002665 SODIUM-D-GLUCOSE COTRA	9.31e+00
4	44	72.1	628	2	067990 METALLOPROTEINASE (FRAGM	9.31e+00
5	44	72.1	1262	5	020771 F54D5.5.	9.31e+00
6	43	70.5	105	4	014421 GLYCOPHORIN M2 (FRAGME	1.56e+01
7	43	70.5	129	4	099737 LAMININ ALPHA 4 CHAIN.	1.56e+01
8	43	70.5	158	5	020474 F46F6.3.	1.56e+01
9	43	70.5	184	5	024743 EGP-1 PRECURSOR.	1.56e+01
10	43	70.5	480	5	027033 MEMBRANE PROTEIN.	1.56e+01
11	43	70.5	1382	11	097523 HGF RECEPTOR PRECURSOR	1.56e+01
12	43	70.5	1382	11	097523 HEPATOCYTE GROWTH FACT	1.56e+01
13	42	68.9	81	4	099483 GLYCOSTYLTRANSFERASE (F	2.58e+01
14	42	68.9	139	5	091281 COSMID F27C1.	2.58e+01
15	42	68.9	156	10	039754 GRPF1.	2.58e+01
16	42	68.9	167	1	059097 167AA LONG HYPOTHETICA	2.58e+01
17	42	68.9	345	14	056987 COAT PROTEIN.	2.58e+01
18	42	68.9	607	5	017112 80 KDA PROTEIN.	2.58e+01
19	42	68.9	631	4	099588 KIAA0078 PROTEIN.	2.58e+01
20	42	68.9	631	4	060216 PROTEIN INVOLVED IN DN	2.58e+01

21	42	68.9	634	11	Q61550 RAD21 HOMOLOG (S. POMB	2.58e+01
22	42	68.9	635	11	P70219 RAD21 HOMOLOG (S.	2.58e+01
23	42	68.9	1211	11	035233 MAF4 (FRAGMENT).	2.58e+01
24	41	67.2	85	5	Q23300 C01F6.8.	4.24e+01
25	41	67.2	140	2	Q45022 XYLX GENE (FRAGMENT).	4.24e+01
26	41	67.2	157	6	062841 CATHALICIDIN PRECURSOR	4.24e+01
27	41	67.2	200	2	067307 HYPOTHETICAL 23.1 KD P	4.24e+01
28	41	67.2	211	2	007446 INSERTION ELEMENT IS14	4.24e+01
29	41	67.2	283	5	026940 HYPOTHETICAL P284 PROT	4.24e+01
30	41	67.2	338	1	028769 HYPOTHETICAL 37.4 KD P	4.24e+01
31	41	67.2	405	10	Q22763 PUTATIVE BZIP-LIKE DNA	4.24e+01
32	41	67.2	482	2	052762 CATALASE (EC 1.11.1.6)	4.24e+01
33	41	67.2	484	2	P77939 CATALASE (EC 1.11.1.6)	4.24e+01
34	41	67.2	493	9	064203 GP10.	4.24e+01
35	41	67.2	625	11	035305 RANK.	4.24e+01
36	41	67.2	662	14	057138 SIMILAR TO HHV6A U76.	4.24e+01
37	41	67.2	919	10	Q24375 ALPHA-GLUCOSIDASE (EC	4.24e+01
38	41	67.2	946	14	Q69139 NUCLEAR ANTIGEN EBNA-3	4.24e+01
39	41	67.2	991	5	018152 T28D6.4.	4.24e+01
40	41	67.2	1186	14	090061 NUCLEAR ANTIGEN-3B (EX	4.24e+01
41	41	67.2	1555	5	044498 F15E5.1 PROTEIN.	4.24e+01
42	40	65.6	330	11	054950 AMP ACTIVATED PROTEIN	6.90e+01
43	40	65.6	586	2	P96351 CTP SYNTHETASE (EC 6.3	6.90e+01
44	40	65.6	702	1	028289 HYPOTHETICAL 78.9 KD P	6.90e+01
45	40	65.6	1078	11	008995 MYELIN TRANSCRIPTION F	6.90e+01

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	374 AA.
ID 042322			
AC 042322;			
DT 01-JAN-1998	(TREMREL. 05, CREATED)		
DT 01-JAN-1998	(TREMREL. 05, LAST SEQUENCE UPDATE)		
DT 01-JUN-1998	(TREMREL. 06, LAST ANNOTATION UPDATE)		
DE D4B DOPAMINE RECEPTOR.			
OS CYPRINUS CARPIO (COMMON CARP).			
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES; GNATHOSTOMATA;			
OC OSTEICHTHYES; ACTINOPTERYGII; CYPRINIFORMES.			
[1]			
RC TISSUE-RETINA;			
RA HIRANO J., ARCHER S.N., DJAMGOZ M.B.A.;			
RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.			
DR EMBL; Y14633; E334823; -			
DR PFAM; PF00001; 7tm_1.			
SQ SEQUENCE 374 AA; 42004 MW; 185EF905 CRC32;			

Query Match 80.3%; Score 49; DB 13; Length 374;
Best Local Similarity 60.0%; Pred. No. 6.39e-01;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 255 EQDPSSEPV 264	PRELIMINARY;	PRT;	344 AA.
Qy 1 QENPDSSEPV 10			

RESULT 2	PRELIMINARY;	PRT;	344 AA.
ID 062447			
AC 062447;			
DT 01-AUG-1998	(TREMREL. 07, CREATED)		
DT 01-AUG-1998	(TREMREL. 07, LAST SEQUENCE UPDATE)		
DT 01-AUG-1998	(TREMREL. 07, LAST ANNOTATION UPDATE)		
DE Y43F4B.1			
OS CAENORHABDITIS ELEGANS.			
OC EUKARYOTA; METAZOA; ACCELOMATES; NEMATODA; SECERNENTIA; RHABDITIDA.			
[1]			
RN SEQUENCE FROM N.A.			
RP MATTHEWS L.;			
RA SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.			
[2]			
RN SEQUENCE FROM N.A.			
RP WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,			

Query Match	72.1%	Score 44;	DB 2;	Length 628;
Best Local Similarity	87.5%;	Pred. No.	9.31e+00;	
Matches	7;	Conservative	1;	Mismatches 0;
		Indels	0;	Gaps 0;

```

Db      98 EXPETSDP 105
        |||::|:|
Qv      2 ENPDSSER 9'

```

```

RESULT 7
ID Q99737 PRELIMINARY; PRT; 129 AA.
AC Q99737
DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE LAMININ ALPHA 4 CHAIN.
GN LAMA4*-1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RA XIAO S., LUX M.L., REEVES R., HUDSON T.J., FLETCHER J.A.;
RL AM. J. PATROL. 0:0-0(0).
DR EMBL; U77706; G1684837; -.
SQ SEQUENCE 129 AA; 13462 MW; 05562347 CRC32;

Query Match 70.5%; Score 43; DB 4; Length 129;
Best Local Similarity 55.6%; Pred.No. 1.56e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 45 QDPPETSEP 53
QY 1 QENPDSSEP 9

RESULT 8
ID Q20474 PRELIMINARY; PRT; 158 AA.
AC Q20474
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE F46F6.3.
OS CAENORHABDITIS ELGANS.
OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RA COTTAGE A.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RL NATURE 368:32-38(1994).
DR EMBL; 250028; G897703; -.
SQ SEQUENCE 158 AA; 18019 MW; C3A480F2 CRC32;

Query Match 70.5%; Score 43; DB 5; Length 158;
Best Local Similarity 55.6%; Pred.No. 1.56e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 83 QHPPEDSEP 91
QY 1 QENPDSSEP 9

RESULT 9
ID Q24743 PRELIMINARY; PRT; 184 AA.
AC Q24743
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)

```

```

DE EGP-1 PRECURSOR.
GN EGP-1.
OS DROSOPHILA VIRILIS (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VIRILIS; TISSUE-SALIVARY GLAND;
RA THUROFF E., STOEVEN S., KRESS H.;
RL MECH. DEV. 37:81-93(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VIRILIS; TISSUE-SALIVARY GLAND;
RX MEDLINE; 90276249.
RA SWIDA U., LUCKA L., KRESS H.;
RL DEVELOPMENT 108:269-280(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-VIRILIS; TISSUE-SALIVARY GLAND;
RX MEDLINE; 90384577.
RA KRESS H., SWIDA U.;
RL NATURWISSENSCHAFTEN 77:317-324(1990).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-VIRILIS; TISSUE-SALIVARY GLAND;
RA JARRIN HENTSCHEL A., THUROFF E., TISCHENDORF B., KRESS H.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN-VIRILIS; TISSUE-SALIVARY GLAND;
RX MEDLINE; 94200049.
RA KRESS H.;
RL CHROMOSOMA 102:734-742(1993).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN-VIRILIS; TISSUE-SALIVARY GLAND;
RX MEDLINE; 90276248.
RA KRESS H., LUCKA L., SWIDA U., THUROFF E., KLEMM U.;
RL DEVELOPMENT 108:261-267(1990).
DR EMBL; Z49942; G887426; -.
KW FLYBASE; FBgn0005594; Dvir\Egpl.
FT SIGNAL.
FT SIGNAL 1 17 POTENTIAL.
FT SIGNAL 1 18 POTENTIAL.
SQ SEQUENCE 184 AA; 20567 MW; C0E0E0FB CRC32;

Query Match 70.5%; Score 43; DB 5; Length 184;
Best Local Similarity 50.0%; Pred.No. 1.56e+01;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 77 BESPEDEEPV 86
QY 1 QENPDSSEP 10

RESULT 10
ID Q27033 PRELIMINARY; PRT; 480 AA.
AC Q27033
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE MEMBRANE PROTEIN.
OS THEILERIA PARVA.
OC EUKARYOTA; PROTOZOA; APICOMPLEXA; SPOROZOA; COCCIDIA; PIROPLASMIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MUGUGA;
RX MEDLINE; 94088665.
RA BAYLIS H.A., ALLSOPP B.A., HALL R., CARRINGTON M.;
RL MOL. BIOCHEM. PARASITOL. 61:171-178(1993).
DR EMBL; L06323; G310893; -.
KW MEMBRANE.
SQ SEQUENCE 480 AA; 52388 MW; 1EDC7723 CRC32;

```

Query Match 70.5%; Score 43; DB 5; Length 480;
 Best Local Similarity 40.0%; Pred. No. 1.56e+01;
 Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 97 QOQPDTPQPI 106
 QY 1 QENPDSSEPV 10

RESULT 11
 ID P97523 PRELIMINARY; PRT; 1382 AA.
 AC P97523;
 DT 01-MAY-1997 (TREMREL. 03, CREATED)
 DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
 DE HGF RECEPTOR PRECURSOR.
 GN C-MET.
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 CC EUTHERIA; RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-LIVER;
 RA WALLENIUS V., RAHET H., SRTIC S., EXBERG S., HELOU K., QIU Y.,
 RA LEVAN G., CARLSSON B., ISAKSSON O., NAKAMURA T., JANSSON J.O.;
 RL MAMM. GENOME 8:661-667(1997).
 DR EMBL; X96786; E238809;
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PFAM; PF00069; pkinase.
 KW SIGNAL.
 FT SIGNAL.
 FT CHAIN 1 24 POTENTIAL.
 FT CHAIN 25 1382 HGF RECEPTOR.
 SQ SEQUENCE 1382 AA; 153940 MW; ED5D6941 CRC32;

Query Match 70.5%; Score 43; DB 11; Length 1382;
 Best Local Similarity 60.0%; Pred. No. 1.56e+01;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 349 QSKPDSAPPM 358
 QY 1 QENPDSSEPV 10

RESULT 12
 ID P97579 PRELIMINARY; PRT; 1382 AA.
 AC P97579;
 DT 01-MAY-1997 (TREMREL. 03, CREATED)
 DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMREL. 06, LAST ANNOTATION UPDATE)
 DE HEPATOCYTE GROWTH FACTOR RECEPTOR.
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 CC EUTHERIA; RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-KIDNEY;
 RA LIU Y., TOLBERT E.M., SUN A.M., DWORKIN L.D.;
 RL AM. J. PHYSIOL. 271:0-0(1996).
 DR EMBL; U65007; G1679660;
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PFAM; PF00069; pkinase.
 SQ SEQUENCE 1382 AA; 153750 MW; 500939CA CRC32;

Query Match 70.5%; Score 43; DB 11; Length 1382;
 Best Local Similarity 60.0%; Pred. No. 1.56e+01;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 349 QSKPDSAPPM 358
 QY 1 QENPDSSEPV 10

RESULT 13
 ID Q99483 PRELIMINARY; PRT; 81 AA.
 AC Q99483;
 DT 01-MAY-1997 (TREMREL. 03, CREATED)
 DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMREL. 05, LAST ANNOTATION UPDATE)
 DE GLYCOSYLTRANSFERASE (FRAGMENT).
 GN ABO.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 CC EUTHERIA; PRIMATES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96437053.
 RA OGASAWARA K., YABE R., UCHIKAWA M., SAITOU N., BANNAI M., NAKATA K.,
 RA TAKENAKA M., FUJISAWA K., ISHIKAWA Y., JUJI T., TOKUNAGA K.;
 RL BLOOD 88:2732-2737(1996).
 DR EMBL; D82841; G1783220;
 KW TRANSFERASE.
 FT NON_TER 1 1
 FT NON_TER 81 81
 SQ SEQUENCE 81 AA; 8346 MW; 32FC1431 CRC32;

Query Match 68.9%; Score 42; DB 4; Length 81;
 Best Local Similarity 60.0%; Pred. No. 2.58e+01;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 72 QEPGPGPEPV 81
 QY 1 QENPDSSEPV 10

RESULT 14
 ID P91281 PRELIMINARY; PRT; 139 AA.
 AC P91281;
 DT 01-MAY-1997 (TREMREL. 03, CREATED)
 DT 01-MAY-1997 (TREMREL. 03, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMREL. 07, LAST ANNOTATION UPDATE)
 DE COSMID F27C1.
 GN F27C1.4.
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; ACOELOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE; 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
 RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAYTON M.,
 RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
 RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
 RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
 RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
 RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWKNEE R., SMAILON N., SMITH A.,
 RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
 RA VAUDIN M., VAUGHAN K., WATERSTON R., WAYSON A., WEINSTOCK L.,
 RA WILKINSON-SPROAT J., WOHLDMAN P.;
 RL NATURE 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WU X., LE T.T.;
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; U80441; G1703603;
 SQ SEQUENCE 139 AA; 14965 MW; 64437070 CRC32;

Query Match 68.9%; Score 42; DB 5; Length 139;
 Best Local Similarity 50.0%; Pred. No. 2.58e+01;

Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 111 HENPDETGP1 120

Qy 1 QENPDSSEPV 10

RESULT 15

ID Q39754 PRELIMINARY; PRT; 156 AA.

AC Q39754;

DT 01-NOV-1996 (TREMELREL. 01, CREATED)

DT 01-NOV-1996 (TREMELREL. 01, LAST SEQUENCE UPDATE)

DT 01-AUG-1998 (TREMELREL. 07, LAST ANNOTATION UPDATE)

DE GRP1.

OS FAGUS SYLVATICA.

OC EURARYOTA; PLANTAE; EMBRYOBIONTA; MAGNOLIOPHYTA; MAGNOLIOPSIDA;

OC HAMAMELIDAE; FAGALES; FAGACEAE.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-SEEDS:

RA NICOLAS C., NICOLAS G., RODRIGUEZ D.;

RL PLANT MOL. BIOL. 36:487-491(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE-SEEDS:

RA NICOLAS C., NICOLAS G., RODRIGUEZ D.;

RL PLANT CELL PHYSIOL. 38:1303-1310(1997).

DR EMBL; X98539; E249668; -.

SQ SEQUENCE 156 AA; 15164 MW; 4B1907FD CRC32;

Query Match

Best Local Similarity 68.9%; Score 42; DB 10; Length 156;

Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 31 EKPDESSEPV 40

Qy 1 QENPDSSEPV 10

Search completed: Thu May 13 15:46:32 1999

Job time : 34 secs.

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 14, 1999, 01:16:20 ; Search time 1151.43 seconds
(without alignments)
7587.411 Million cell updates/sec

Title: US-09-040-485-1
Perfect score: 2442
Sequence: 1 CGGAGCTTGAAGGACACAA.....GGTTAACTTTAAATATTTT 2442

Scoring table: IDENTITY_NUC

Searched: 808301 seqs, 1788773984 residues

Database : GenEmbl.*

- 1: gb_bal.*
- 2: gb_ba2.*
- 3: gb_in.*
- 4: gb_cm.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl1.*
- 9: gb_pl2.*
- 10: gb_pr1.*
- 11: gb_pr2.*
- 12: gb_pr3.*
- 13: gb_ro.*
- 14: gb_st.*
- 15: gb_sy.*
- 16: gb_un.*
- 17: gb_v1.*
- 18: gb_vtg.*
- 19: em_ba.*
- 20: em_fun.*
- 21: em_hum1.*
- 22: em_hum2.*
- 23: em_in.*
- 24: em_om.*
- 25: em_or.*
- 26: em_ov.*
- 27: em_pat.*
- 28: em_ph.*
- 29: em_pl.*
- 30: em_ro.*
- 31: em_sy.*
- 32: em_un.*
- 33: em_v1.*
- 34: em_vtg.*
- 35: em_sts.*
- 36: gb_bal.*
- 37: gb_ba2.*
- 38: gb_pl1.*
- 39: gb_pl2.*
- 40: gb_pr1.*
- 41: gb_pr2.*
- 42: gb_pr3.*
- 43: gb_sts.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description

1	825.4	33.8	2449	10	HSU03109	U03109 Human aspar
2	825.4	33.8	2449	40	HSU03109	U03109 Human aspar
3	823.8	33.7	2324	11	S83325	S83325 aspartyl(as
4	823.8	33.7	2324	41	S83325	S83325 aspartyl(as
5	342.6	14.0	2739	4	BOVASBHY	M91213 Bos taurus
6	183	7.5	1777	4	CFU38414	U38414 Canis famli
7	97.4	4.0	2069	27	E10126	E10126 DNA encodin
8	97.4	4.0	3399	27	E10126	E10126 DNA encodin
9	95.2	3.9	1686	27	E08995	E08995 DNA encodin
10	87.6	3.6	380	43	G23118	G23118 human STS W
11	84.4	3.5	86	6	AR003317	AR003317 Sequence
12	79.4	3.3	112930	17	HSGEND	X64346 Herpesvirus
13	79.4	3.3	43658	17	HSV3PRGEN	M86409 Herpesvirus
14	79.4	3.3	3720	17	S76368	S76368 ORF 5' of E
15	77.6	3.2	188324	18	AC005505	AC005505 *** SEQUE
16	71	2.9	116884	18	PFMALIPI	AL031744 Plasmodiu
17	69.6	2.9	1324	3	PFRESAR1	X05182 P.falciparu
18	69.6	2.9	4591	6	A00661	A00661 P.falciparu
19	68.6	2.8	996	27	E08996	E08996 DNA encodin
20	68.6	2.8	585	27	E08997	E08997 DNA encodin
21	68.2	2.8	5420	3	AF056936	AF056936 Plasmodiu
22	68	2.8	7218	6	I66494	I66494 Sequence 14
23	67.8	2.8	2384	3	PFAGAR	J03998 P.falciparu
24	67.4	2.8	3777	9	SPBC215	AL033534 S.pombe c
25	67.4	2.8	3777	39	SPBC215	AL033534 S.pombe c
26	67	2.7	8574	3	AF057019	AF057019 Dictyoste
27	66.8	2.7	914	3	PFRESA	X55124 P.falciparu
28	65.6	2.7	4050	3	DDPROKING	Z37981 D.discoideu
29	65	2.7	4739	3	PFAGLURPA	M59706 P.falciparu
30	65	2.7	2349	6	A00764	A00764 Synthetic p
31	64.2	2.6	42141	3	CXY54G9A	AL032648 Caenorhab
32	64.2	2.6	300197	18	CEY54G9	Z98869 Caenorhabdi
33	64	2.6	12029	3	AE001373	AE001373 Plasmodiu
34	64	2.6	334	8	HVC39SAT	Z50100 H.vulgare G
35	64	2.6	334	38	HVC39SAT	Z50100 H.vulgare G
36	62	2.5	2941	5	XLU85969	U85969 Xenopus lae
37	62	2.5	397	43	G37798	G37798 garp plasm
38	61.6	2.5	3763	5	PMU19361	U19361 Petromyzon
39	61	2.5	3337	6	I23337	I23337 Sequence 1
40	59.6	2.4	15421	3	PFCOMPIRA	X95275 P.falciparu
41	59.6	2.4	631	3	PFPPPK51	X53026 P.falciparu
42	59.6	2.4	2967	3	PFSC03115	AL021887 Plasmodiu
43	59.6	2.4	3010	5	XLU85970	U85970 Xenopus lae
44	59.6	2.4	40873	18	CEH04109	Z92848 Caenorhabdi
45	59	2.4	15148	3	AE001393	AE001393 Plasmodiu

ALIGNMENTS

RESULT 1	HSU03109	HSU03109	2449 bp	MRNA	PRI	30-NOV-1995
LOCUS	Human aspartyl	beta-hydroxylase	mrna	complete cds.		
DEFINITION	U03109					
ACCESSION	g458031					
NID						
KEYWORDS	human.					
SOURCE	Homo sapiens					
ORGANISM	Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
REFERENCE	1 (bases 1 to 2249)					
AUTHORS	Korloeth,F., Gieffers,C. and Frey,J.					
TITLE	Cloning and characterization of the human gene encoding aspartyl beta-hydroxylase					
JOURNAL	Gene 150 (2), 395-399 (1994)					
MEDLINE	95121937					
REFERENCE	2 (bases 1 to 2449)					
AUTHORS	Korloeth,F.					
TITLE	Direct Submission					
JOURNAL	Submitted (03-NOV-1993) Korloeth F., Fakultat fuer Chemie-Biochemie II, Universitaet Bielefeld, Universitaetsstrasse 25, Bielefeld, 33615, Germany					

Accession	Length	Source	Organism	Gene	Protein	Notes
AGTTTCACAAGACTGTAATCAGGATATGGAAGAGATGATGTTCTGAGCAGGAGAAATCCAG	604	Human	Homo sapiens			
AGTTTCACAAGACTGTAATCAGGATATGGAAGAGATGATGTTCTGAGCAGGAGAAATCCAG	786	Human	Homo sapiens			
TTCCAGTGAACCAAGTAGTAGAAGATGGAAGATGTCACCATGATACACAGATGATGTRACAT	664	Human	Homo sapiens			
TTCCAGTGAACCAAGTAGTAGAAGATGGAAGATGTCACCATGATACACAGATGATGTRACAT	846	Human	Homo sapiens			
CCAAGTCTATGAGGAACAAGCAGTATATCAACTCTAGAAAAATGAAGGGATAGAAATCA	724	Human	Homo sapiens			
CCAAGTCTATGAGGAACAAGCAGTATATCAACTCTAGAAAAATGAAGGGATAGAAATCA	906	Human	Homo sapiens			
AGAAGTAACTGCTCCCTCGAGGATAATCCTCTAGAAAGATTCACAGGTAATTTGTAGAAG	784	Human	Homo sapiens			
AGAAGTAACTGCTCCCTCGAGGATAATCCTCTAGAAAGATTCACAGGTAATTTGTAGAAG	966	Human	Homo sapiens			
AGTAGCATTTTTCTGTGGGAGACAGCAGGAAGTACCACAGATACCTTAAAG	839	Human	Homo sapiens			
AGTAGCATTTTTCTGTGGGAGACAGCAGGAAGTACCACAGATACCTTAAAG	1021	Human	Homo sapiens			
HSU03109	2449 bp	mrna	PRI	30-NOV-1995		
Human aspartyl beta-hydroxylase mRNA, complete cds.						
U03109						
9458031						
human.						
Eukaryotes; mitochondria eukaryotes; Metazoa; Chordata;						
Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
1 (bases 1 to 2449)						
Korloth, F., Gieffers, C. and Frey, J.						
Cloning and characterization of the human gene encoding aspartyl						
beta-hydroxylase						
Gene 150 (2), 395-399 (1994)						
95121937						
2 (bases 1 to 2449)						
Korloth, F.						
Direct Submission						
Submitted (03-NOV-1993)						
II, Universitaet Bielefeld, Universitaetsstrasse 25, Bielefeld,						
33615, Germany						
Location/Qualifiers						
1. .2449						
/organism="Homo sapiens"						
/db_xref="taxon:9606"						
/clone="AS-5"						
/clone_lib="MG63-2AP"						
/cell_line="mg63"						
/cell_type="osteosarcoma"						
1. .77						
78. .2351						
/codon_start=1						
/function="hydroxylation of aspartyl and asparaginyl						
residues"						
/product="aspartyl beta-hydroxylase"						
/db_xref="PID:g458032"						
/translation="MAQRNKAASSNSSSSGGSGSTAGSSSPGARRETHKGHHNG						
AKGLVLSRFTTWFVIALGVTVAVWFVDLYEVYGLKGLGYDADGDFD						
RVGLLGGSTTSEPAVPEAEAPTEPEQVPAEPQNTIDEAKQIQSLLLHVEV						
AEHVGEQLQEDGPTGPQEDDEFLATVDVDFETLEPVSHEETEHSHVHEV						
SODCNQDMEEMSEQENPDSEPPVYDESLHDDTDVYQYEEQAVPELENGE						
TEVTPADPNVEDSQVIEVSIPFVEEQEVPETNRKTDDEPQAKRVKKPKLL						
NRFDKTIKALDAAELKRGKIEPAVNAFLKRVKPSQSPARYTKAQCEDDLAEK						
RSFVLRGATITQEVASLPVADLLLSLRKRRSDROQFLGHMRGSLTLQRLV						
PNDSLKGLDGVYLLIGDNDNAKVYEVLSVTPNDGFAKHVYGLKAKNQIAES						
PYLKGIIESGPDGDDGRFYELGDGMQRVGNKEAYKVELGHRGHFASVQW						
VNGLKAQPCGPKETGYQLVSLERNWKLIRDEGLAVMDKAGUFLPEDENLR						
QFTLVQGRNRNENAKGPKTTLLEKFPETGTCRGQIKYSIMHPHTVMPHT						
NCRLRHMLGVIPEKRGKIRCANETKTWEKGKVLIFDSSFHEVMQDASSFR						
LPIFLPIV						

QY	305	AGCAGAACCCCAAGATATCGAAGATGAACAAAGAACAAATTCAGTCCCTTCTCCATG	364
Db	421	AGCAGAACCCCAAGATATCGAAGATGAACAAAGAACAAATTCAGTCCCTTCTCCATG	480
QY	365	AAATGGTACAGCGACAACATGTTGAGGGAAGACTTTGCAACGAAGAGATGGACCCACAG	424
Db	481	AAATGGTACAGCGACAACATGTTGAGGGAAGACTTTGCAACGAAGAGATGGACCCACAG	540
QY	425	GAGAACCAACAAAGAGGATGATGAGTTCTTATGGCGACTGATGATGATGATGATG	484
Db	541	GAGAACCAACAAAGAGGATGATGAGTTCTTATGGCGACTGATGATGATGATGATG	600
QY	485	AGACCTTGGAACTGGAATATCTCATGAAGAACCGAGCATAGTTACCGGTGAAGAGA	544
Db	601	AGACCTTGGAACTGGAATATCTCATGAAGAACCGAGCATAGTTACCGGTGAAGAGA	660
QY	545	CAGTTTCACAAGACTGTAATCAGGATATGAAGAGATGATGCTGACGAGGAATCCAG	604
Db	561	CAGTTTCACAAGACTGTAATCAGGATATGAAGAGATGATGCTGACGAGGAATCCAG	720
QY	605	ATTCCAGTGAACCACTAGTAGAAGATGAAGATGACCATGATACAGATGATGTAACAT	664
Db	721	ATTCCAGTGAACCACTAGTAGAAGATGAAGATGACCATGATACAGATGATGTAACAT	780
QY	565	ACCAAGTCTATGAGGACAAGCAGTATATGAACCTCTAGAAAATGAAGGATAGAAATCA	724
Db	781	ACCAAGTCTATGAGGACAAGCAGTATATGAACCTCTAGAAAATGAAGGATAGAAATCA	840
QY	725	CAGAAGTAACCTGCTCCCTGAGGATATCCTCTAGAAGATTCACAGGTAATGTAAGAG	784
Db	841	CAGAAGTAACCTGCTCCCTGAGGATATCCTCTAGAAGATTCACAGGTAATGTAAGAG	900
QY	785	AAGTAAGCATTTTTCTCTGTGGAAGAACGACGAGGAAGTACCACAGATATCTTAAAG	839
Db	901	AAGTAAGCATTTTTCTCTGTGGAAGAACGACGAGGAAGTACCACAGATATCTTAAAG	955
RESULT	4		
LOCUS	S83325	2324 bp mRNA	27-MAR-1997
DEFINITION	aspartyl(asparaginyl)beta-hydroxylase [human, hepatoblastoma cell line HepG2, mRNA Partial, 2324 nt].	PRI	
ACCESSION	S83325		
NID	g1911651		
KEYWORDS	human hepatoblastoma cell line HepG2.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
REFERENCE	1 (bases 1 to 2324)		
AUTHORS	Lavalisere, L., Jia, S., Nishiyama, M., de la Monte, S., Stern, A.M., Wands, J.R. and Friedman, P.A.		
TITLE	Overexpression of human aspartyl(asparaginyl)beta-hydroxylase in hepatocellular carcinoma and cholangiocarcinoma		
JOURNAL	J. Clin. Invest. 98 (6), 1313-1323 (1996)		
MEDLINE	96420598		
REMARK	GenBank staff at the National Library of Medicine created this entry [NCBI gibsseq 179962] from the original journal article. This sequence comes from Fig. 4.		
COMMENT	Authors note differences between this sequence and that of F. Koriolth, C. Gleffers, and J. Frey: Gene 150 (2), 395-399 (1994), GenBank U03109.		
FEATURES	Location/Qualifiers		
source	1..2324		
	/organism="Homo sapiens"		
gene	/db_xref="taxon:9606"		
	12..2288		
	/gene="aspartyl(asparaginyl)beta-hydroxylase, HAAH"		
CDS	12..2288		
	/gene="aspartyl(asparaginyl)beta-hydroxylase, HAAH"		
	/note="This sequence comes from Fig. 4; HAAH"		
	/codon_start=1		
	/product="aspartyl(asparaginyl)beta-hydroxylase"		

/db_xref="PID:g1911652"	
/translation="MAQRNNAKSSSSSSSGSGSGSTAGSSSPGARRETKHGHKHNK RKGLSGSTFTWMFIALLGVSVAVWFDVLVEVLKGLIYDADGDGDFVDVH AKVLLGLKERSTSEPAVPPEAEAPHTPEQVPEAEQONIEDAEKEQIQSLLEHVV AEHVGEDELDQEDGPTGEQOEDDEFLMATDVIDDRETLEPEVSEHTEHYSVEEY SQCNQDEMMSESNPQSDSEPVVEDERLJHDTDDYTVQVIEEQVPELENGEIEI TEYATPPENPDSQIVIEEVSIFPVEEQEVPPTNRTDDPEQAKVKKKKPKLL NKFDPTIKAPLDAEKLRRKGLTEEVNAFKELVRYKYPQSPRAYKQACEDDLAKR RSENLKDLGTQYEVASLPDPAADLLKSSRRSQOFLGHMRGSLTLQRLVOLF PNDTSLKNDLGVGLINDNDNAKKVYEVLSVTPNDGFAKVHYGTLKRAQNIASEI PYLKEGESDPGTDGGRVYFHLGDAMQVGNKEAYKWELGHKRGHFASVWORSLYN VNGLKAPWMTPKETGTEGLLSERLNKLRIDEGLAVMQDKAKYGLPEDENUREKGD WSQTLTQQQRRNNEKCAKPKTCTLLEKFPETTGRCQGIKYSIMHPGTHVWPTGP TNCRLRMHLGLVLPKGEKGRICANETRTWEEGKVLIFDDSEFEHVQDASFSRLIFIV DVWHPELTPOORSLPAI"	
BASE COUNT	753 a 463 c 625 g 483 t
ORIGIN	
Query Match	33.7%; Score 823.8; DB 41; Length 2324;
Best Local Similarity	99.3%; Pred. No. 2.8e-132;
Matches	826; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
Qy	5 AGCTTGAGGACACAAAGATGGGAGGAAGCGGACCTCTCAGGAACCTCATCTTTCACGT 64
Db	121 AGCATGGAGGACACAAAGATGGGAGGAAGCGGACCTCTCGGGAACCTCATCTTTCACGT 180
Qy	65 GGTATTATGGTGATTCGATTGCTGGCGCTGCTGGACATCTGTAGCTGCTGTTGGTTTCATC 124
Db	181 GGTATTATGGTGATTCGATTGCTGGCGCTGCTGGACATCTGTAGCTGCTGTTGGTTTCATC 240
Qy	125 TTGTTGACTATGAGGAAGTTCTTAGGAAACTAGGAATCTATGATCGTGGTGGTATGGAG 184
Db	241 TTGTTGACTATGAGGAAGTTCTTAGGAAACTAGGAATCTATGATCGTGGTGGTATGGAG 300
Qy	185 ATTTTGATGGATGATGCGCAAGTTTATTAGACHTTAAGAGAGATCTACTTTCAGAGC 244
Db	301 ATTTTGATGGATGATGCGCAAGTTTATTAGACHTTAAGAGAGATCTACTTTCAGAGC 360
Qy	245 CAGCAGTCCGCCAGAGAGAGCTGAGCCACACACTGAGCCGAGCCGAGGACAGTTCCTCTGG 304
Db	361 CAGCAGTCCGCCAGAGAGAGCTGAGCCACACACTGAGCCGAGGAGGAGCTTCTCTGG 420
Qy	305 AGGCAGAACCCCAAGATATCGAAGATGAAGCAAGAACAAATTCAGTCCCTTCTCCATG 364
Db	421 AGGCAGAACCCCAAGATATCGAAGATGAAGCAAGAACAAATTCAGTCCCTTCTCCATG 480
Qy	365 AAATGGTACAGCGACAGACATGTTTGAGGGAAGACACTTGCACAGAAAGATGGAGCCACAG 424
Db	481 AAATGGTACAGCGACAGACATGTTTGAGGGAAGACACTTGCACAGAAAGATGGAGCCACAG 540
Qy	425 GAGAACACACACAGAGGATGATGAGTTCTTATGGCGACTGATGTAGATGATAGATTG 484
Db	541 GAGAACACACACAGAGGATGATGAGTTCTTATGGCGACTGATGTAGATGATAGATTG 600
Qy	485 AGACCCCTGGAACCTGGAAGTATCTCATGAAGAACCCGAGCATAGTTACACAGTGGGAAGAGA 544
Db	601 AGACCCCTGGAACCTGGAAGTATCTCATGAAGAACCCGAGCATAGTTACACAGTGGGAAGAGA 660
Qy	545 CAGTTTCACAAGACTGTAATCAGGATATGGAAGAGATGATGTTGAGCAGGAGAAATCCAG 604
Db	661 CAGTTTCACAAGACTGTAATCAGGATATGGAAGAGATGATGTTGAGCAGGAGAAATCCAG 720
Qy	605 ATTCCAGTGACACAGTAGTAGAGAGATGAAGATGACACATGATACAGATGATGTACAT 664
Db	721 ATTCCAGTGACACAGTAGTAGAGATGAAGATGACACATGATACAGATGATGTACAT 780
Qy	665 ACCAAGTCTATGAGGAACAAGCAGTATATCAACTCTAGAAAAATGAAGGATAGAAATCA 724
Db	781 ACCAAGTCTATGAGGAACAAGCAGTATATCAACTCTAGAAAAATGAAGGATAGAAATCA 840
Qy	725 CAGAAGTAACCTGCTCCCCCTGAGGATTAATCCTGTAGAAGATTACACAGGTAATGTGAAG 784

MEDLINE 96107245
 REFERENCE 2 (bases 1 to 1777)
 AUTHORS Jones,L.R., Zhang,L., Sanborn,K., Jorgensen,A.O. and Kelley,J.
 TITLE Direct Submission
 JOURNAL Submitted (12-OCT-1995) Larry R. Jones, Medicine, Krannert Inst.
 Cardiology, 1111 W. 10 Street, Indianapolis, IN 46202, USA
 FEATURES Location/Qualifiers
 source 1..1777
 /organism="Canis familiaris"
 /db_xref="taxon:9615"
 /tissue_type="heart"
 89..721
 /note="junctin"
 /codon_start=1
 /function="calcineurin binding protein"
 /product="junctional sarcoplasmic reticulum protein"
 /db_xref="pid:g1163913"
 /translation="MAPETKHGKNGKGLSGSSFTWFVIALGLVYTSVAVWF
 DLVDYEVGLKLVGDGDFDDVDKVLLEGGPGVAKRKTAKVKELTKELKKE
 KETESRKNENRERKKEDEKDKIADADISRKESPKGKDKREKNVGLDKS
 AKAKESRKSNTVKAASSASTASRDKDDTKEGKTSKTHSAKGNQKRKN"
 polyA_site 1777
 /note="44 A nucleotides"
 BASE COUNT 552 a 331 c 363 g 531 t
 ORIGIN

Query Match 7.5%; Score 183; DB 4; Length 1777;
 Best Local Similarity 90.7%; Pred. No. 1.6e-22;
 Matches 195; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
 QY 5 AGCTTGAGGACACAAATGGGAGGAGGCGACCTCAGGAACCTTCATCTTCACGT 64
 DB 105 AGCATGAGGACACAAATGGGAGGAGGAGGACCTTCGGAAGTTCATTTTCACAT 164
 QY 65 GGTATTGATGCTATGCTGGCGCTCTGGACATCTGTAGCTGTCTTTGGTTTGTATC 124
 DB 165 GGTATTGATGCTATGCTGGCGCTCTGGACATCTGTAGCTGTCTTTGGTTTGTATC 224
 QY 125 TTGTTGACTATGAGGAAGTTCTTAGGAAACTAGGAATCTATGATGCTGATGGTGGAG 184
 DB 225 TTGTTGATTGAGGAAGTTCTTAGGAAACTAGGAGTCTATGATGCTGATGGTGGAG 284
 QY 185 ATTTTGATGCTGATGCCCAAGTTTATTAGGA 219
 DB 285 ATTTTGATGCTGATGCCCAAGTTTATTAGGA 319

RESULT 7
 ID E10125 standard; RNA; UNC; 2069 BP.
 XX AC
 XX AC
 XX NI
 XX d1108462
 DT 08-OCT-1997 (Rel. 52, Created)
 DT 08-OCT-1997 (Rel. 52, Last updated, Version 1)
 XX
 DE DNA encoding an immunogenicity protein.
 XX
 KW JP 1995284392-A/1.
 XX
 OS unidentified
 OC unclassified.
 XX
 RN [1]
 RP 1-2069
 RA Doi H., Nagakuchi Y., Tanaka Y., Fujisaki Y.;
 RT "GENE CLONE MANIFESTING CHICKEN LEUCOCYTOZOON IMMUNOGENIC PROTEIN AND
 RT GENE RECOMBINANT VACCINE AGAINST CHICKEN LEUCOCYTOZOON";
 RL Patent number JP 1995284392-A/1. 31-OCT-1995.
 RL DOUBUTSUYOU SEIBUTSUGAKUTEKI SEIZAI KYOKAI, KITASATO INST:THE.

XX OS Leucocytozon caulleryi
 CC PN JP 1995284392-A/1
 CC PD 31-OCT-1995
 CC PF 19-APR-1994 JP 1994080643
 CC PI DOI HIROHITO, NAGAKUCHI YOSHIO, TANAKA YOSHIO, FUJISAKI YUJIRO
 CC PC C12N15/09,A61K39/015,C12P21/02;
 CC CC strandedness: Double;
 CC CC topology: Linear;
 CC CC key Location/Qualifiers
 CC FH source 1..2069
 CC FT /organism="Leucocytozon caulleryi"
 CC FT /strain="shizuoka"
 CC FT 1..25
 CC FT misc_feature /note="EcoRI adaptor"
 CC FT misc_feature 2045..2069
 CC FT /note="EcoRI adaptor"
 CC FT misc_feature 25..2046
 CC FT /product="immunogenicity protein"
 XX Key Location/Qualifiers
 FH source 1..2069
 FT /organism="unidentified"
 FT
 XX Sequence 2069 BP; 1216 A; 167 C; 443 G; 243 T; 0 other;

Query Match 4.0%; Score 97.4; DB 27; Length 2069;
 Best Local Similarity 46.3%; Pred. No. 7e-08;
 Matches 320; Conservative 0; Mismatches 371; Indels 0; Gaps 0;
 QY 147 AGGAAACTAGGAATCTATGATGCTGTGATGGTGGAGATTTTGTGATGGATGCCAA 206
 DB 1053 AGAAAAGAAAGTACACATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1112
 QY 207 AGTTTATTAGGACTTAAAGAGAGATCTACTTCAGAGCCAGCAGTCCCGCCAGAGAGGC 266
 DB 1113 AGAAAAATTAATACATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1172
 QY 267 TGACCCACACTGAGCCCGAGGAGGAGGTTCTTGTGGAGCAGACCCCAAGATATCGA 326
 DB 1173 AGTAACACATGAAG 1232
 QY 327 AGATGAAGCAAG 386
 DB 1233 ACATGAAG 1292
 QY 387 TGAGGAG 446
 DB 1293 AGAAG 1352
 QY 447 TGAGTTTCTTATGCGGACTGATGTAGATGATGATGATGATGATGATGATGATGATG 506
 DB 1353 AGAAG 1412
 QY 507 TCATGAAG 566
 DB 1413 ACATGAAG 1472
 QY 567 GGATATGAAGAGAGAGATGATGCTGTAGCAGGAGAGAGAGAGAGAGAGAGAGAGAG 626
 DB 1473 ACATGAAG 1532
 QY 627 AGATGAAG 686
 DB 1533 AGAAG 1592
 QY 687 AGTATGAAG 746
 DB 1593 AGAAG 1652

*POLYPEPTIDE INDUCING IMMUNITY AGAINST LEUCOCYTOZOON PROTOZOAN AND
RECOMBINANT DNA MOLECULE CODING THE POLYPEPTIDE*;
Patent number JP 1995089995-A/1, 04-APR-1995
DOUBUTSUYOU SEIBUTSUGAKUTEKI SEIZAI KYOKAI, NIISUSEIKEN KK.

OS Leucocytozoon caulleryi
CC PN JP 1995089995-A/1
CC PD 04-APR-1995
CC PF 10-SEP-1993 JP 1993226078
CC PI KATO ATSUSHI, ONAGA HIROSHI, UEDA SUSUMU
CC PC C07K14/44;A61K39/015,C12N1/21,C12N15/09,C12P21/02,G01N33/53,
CC PC G01N33/569,
CC PC (C12N1/21,C12R1:19),(C12P21/02,C12R1:19);
CC CC strandedness: Double;
CC CC topology: Linear;
CC FH Key Location/Qualifiers

CC FT source 1..1686
CC FT /organism="Leucocytozoon caulleryi"
CC FT /note="mRNA is derived from 2nd shizont"
CC FT mat_peptide 1..1686
CC FT /product="a protein involved in Leucocytozoon
CC FT phylaxis"

XX Key Location/Qualifiers
FH source 1..1686
FT /organism="unidentified"
XX
SQ Sequence 1686 BP; 915 A; 129 C; 382 G; 260 T; 0 other;

Query Match 3.9%; Score 95.2; DB 27; Length 1686;
Best Local Similarity 46.9%; Pred. No. 1.7e-07;
Matches 335; Conservative 0; Mismatches 373; Indels 6; Gaps 1;

QY 137 AGGAAGTCTTAGGAAACTAGGAATCTATGATGCTGATGGTGTGATGAGATTTTGTGTCG 196
DB 125 AAGAGACACAG 184
QY 197 ATGATGCCAAAGTTTATTAGGACTTAAAGAGAGATCTACTTCAGAGCCAGCAGTCCCGC 256
DB 185 AACAGAACAGATGAAGAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 244
QY 257 CAGAGAGCGCTGAGCCACACTGAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 316
DB 245 AAG 304
QY 317 AGAATATCGAAGATGAAG 376
DB 305 ATGAG 358
QY 377 CAGAACATGTTGAGGAG 436
DB 359 ATGAG 418
QY 437 AAGAGGATGATGAGTTCTTATGGCGACTGATGATGATGATGATGATGATGATGATGATG 496
DB 419 ATGAAGATGAACAAATGAAG 478
QY 497 CTGAGTATCTCATGAAG 556
DB 479 ATGAAG 538
QY 557 ACTGTAATCAGGATGAG 616
DB 539 AACAGATGAG 598
QY 617 CAGTAGTAG 676
DB 599 AAG 658
QY 677 AGGAG 736

DB 659 AAGTATATCTAGAAAAAGAACAGAACATTTAAAGAAATTAGAAAAAGAGTTTGAAG 718
QY 737 CTCCTCCTGAGGATATCTGTAGAGAGATTCACAGTAAATTGTAGAGAGAGTAAAGCATTT 796
DB 719 AAAAGTAGAAGATGAACATAAAGATGATAAGAAAGAGACAAAAATAAAAAAGCATGAAG 778
QY 797 TTCTCTGTGAAGAACAGCAGGAGAGTACCACCATGATCTTAAAGCTTCAAAAAACA 850
DB 779 TTGAAGAACAGCAGAGAAATAGAAATAAAGGAGATCAGAAGAAATAGATGAGA 832

RESULT 10

G23118/c 380 bp DNA STS 31-MAY-1996
LOCUS human STS WI-11767.
DEFINITION G23118
ACCESSION
NID g1343444
KEYWORDS STS sequence; primer; sequence tagged site.
SOURCE human STSs derived from sequences in dbEST and the Unigene collection.

ORGANISM Homo sapiens
Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 380)
AUTHORS Hudson,T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically
Mapped STS
JOURNAL Unpublished (1995)
COMMENT

Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: ATTCTGTATTTTATGCACTCAACA
Primer B: ATTCCAAATATCTGAATTAATGAC
STS size: 125
PCR Profile:

Presoak:
Denaturation:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:
Template: 10 ng
Primer: each 5 pm
dNTPs: each 4 nM
Taq Polymerase: 0.025 units/ul
Total Vol: 20 ul

Buffer:

MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCL: 10 mM
pH: 9.3

Derived from dbEST (genbank accession R08579).

FEATURES
source

1..380
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="422.5 cR from top of Chr8 linkage group"

STS

primer_bind 7..131

primer_bind 7..31

complement(107..131)

BASE COUNT 116 a 73 c 64 g 122 t 5 others

ORIGIN

exon
PAGTTLHTVTCREGYAKRPVQVTCVNGNWTVPVKCKOKKCKCSTPPQDLLNGRYVTGNLY
YGSVITYTCNGSYSLIGSTTSACLLKRGGRVDMWTTPPICDKKCKPPQIANGHTTN
VKDYTYLDYTYTSCNDTTLKLTGTPSSKLACSETGSPVNGETKCEFFICKLPQVANA
YVEVRKSAATSMOYLHINVKYKGFMLYGETPTNTCHGVWSPAIPCEMKISSPKGDMG
INSNEDNSTPSGAECACGSPNPISS"
<10912..11773
/gene="mCCPH"
/number=1
10912..11994
/gene="mCCPH"
10912..12013
/gene="sCCPH"
10912..11994
/gene="mCCPH"
/note="ORF 04a"
/codon_start=1
/evidence=experimental
/product="complement control protein homologue"
/db_xref="PID:g60324"
/db_xref="SWISS-PROT:P01016"
translation="MYTLHYICLVLSGVYFVWTLSCPTRNOYVSVKYNVNLTSQPY
PAGTTLHTVTCREGYAKRPVQVTCVNGNWTVPVKCKOKKCKCSTPPQDLLNGRYVTGNLY
YGSVITYTCNGSYSLIGSTTSACLLKRGGRVDMWTTPPICDKKCKPPQIANGHTTN
VKDYTYLDYTYTSCNDTTLKLTGTPSSKLACSETGSPVNGETKCEFFICKLPQVANA
YVEVRKSAATSMOYLHINVKYKGFMLYGETPTNTCHGVWSPAIPCEMKISSPKGDMG
INSNEDNSTPSGRICNGCTTSMPTQTITITARTYSTHIYPTGTXYLKPGLVLIIL
TTSPIIGIILITGVCLHRCVCMGSGQ"
11774..11966
/gene="mCCPH"
/number=1
11967..>12013
/gene="sCCPH"
/number=2
complement(11995..12285)
/note="ORF 05; KFLF1"
/codon_start=1
/db_xref="PID:g60326"
/db_xref="SWISS-PROT:P24912"
translation="MTEIFLFLKAKHLLOAEENNYNDPVYEIQIPLSECNRTT
SLNSIYDDVCYPEDMDYDDCGAYTRLDASVDYGLRTCGNDGSHIYEIG"
12384..15970
/note="ORF 06; KRF1"
/codon_start=1
/product="major ssDNA-binding protein"
/db_xref="PID:g60327"
/db_xref="SWISS-PROT:P24910"
translation="MATKAQPSADNLGRAPVEPCGYIYVYPRKGFPPKEASLAGNK
NVGASLPLLSLDTVESNFSFNVAHKKIDMTLLVRYSAHYREAIVFNTDLFE
PIFVPGDLILCDARSDFGTVNVPRTDLRDTVDIKDLYAPFYSEDSCFMAVVYTEG
FKERLYGNLPIITAOGLKVOINGREAVKIPLYDEDLFSKSHLPRYIPSVKYLH
DSVETSIAQALRIDVESVIRASEKQIQDQYKLAKIVNSKDFSLQSQCDASAFMV
IDCIAELAMSGISLFLAPQDPCAVLDYTSWPIFETAETEDRKAIDQNMAMSVH
VYTHLSTNSVLYLTINKQPSKSEQNVNTFMQHLAYADAQOREGEPAFSG
APKFSGGTYTLHLAASFPHLIARNYVMQFCQHKQSTNANYSPVQVYGTAAAS
DLCELCQTCASGCIHTFYRLKRPPLVGLSORRDPVYVTVGSQYNDLDMGNFAT
FREKEDEAONAESEKTYWQIQNVVEKLSMGVTEGTVGSLIITDQSLKTFRDI
DNVYDSEVKKMNLKYNKNNIFREIKTVHVLHYCCNVFWQAPCAMFLFYKSLVA
LIQDICIPIANTISODNSIGMPSSEWLKHQYQITWTFKSCDLRGLVLTSEKRIHV
TDFCFDLNLSLQSDGIVPMKQVRLAKALLTPKTIKRNIVFSNMSMETIQSG
FIKAKTKDSYIVTPGYKFLNSLHKVMFNPAKISALYLWHTFSOKKOLPVLPGISRE
NMVELANVETSSKHMDDMVDIIPITLLTYAKVRLANTILRTCGQTFVATLOCL
LPTQITISATYPHVLLDQSIMSDHYLSSIKDRHALVQVTLKEDIAVQKQPIVY
VPLVYVYTGINGNTQIFCCQNLGTFMGRGVDRLNIPDSTFRFRONNSYMRHRHVF
TPMVAHLVKKNSLNNLTFEVETIRKNQONIFEDKNLNIIDFNVDYLEVLVGLGDSCE
ITEDDLQFYLGEEYIMSEIWSRFQIITDSGAPMSVENVTYVLCGNKQCECKFEFVG
EEOQSCVPPQTEEFAPQATLTLAASRRKRTSILSDIDL"
15977..18016
/note="ORF 07; KRF2"
/codon_start=1
/product="transport protein"
/db_xref="PID:g60328"
/db_xref="SWISS-PROT:P24911"
translation="MAQLAANYQIYGLTLDVSLITFDPAHINWKIVTNTKINK

CDS
18006..20432
/note="ORF 08; KCRF1"
/codon_start=1
/product="glycoprotein B"
/db_xref="PID:g60329"
/db_xref="SWISS-PROT:P24905"
translation="MVPNKHLLLIILSFSTAGQTTPTTAVAKNKTAIQYFYRYR
CSAATTELFRLDRTCTPCTEDKHVKIGILLVYKKNIIVPIYFKVRYRKKITTSNRIF
NGWTRGVATNKLWELSAVPKYEIDIMDKTYQCHNCMOIEVNGMRYVDRGNKKT
VDLKPVDGLTGATIRYISQPKVFAADPGHMLGTYRTVNCNCEIVDMFARSADPYTFV
TALGDTVEPCDNDSCPNATDLSVOIDLNIHTVVDIGNRATSOQHKKRIFAHTLD
YSVSEAVNKSASVCSVFWKSFQRAIQIOTDEHDLTYHTANEITAGFTVKEPLANFTS
DYNCLMTINTTLEDKIARVNNHTPTAEYQTEGGMLVWQPLIAIELEAMELEA
TTSPTYAPATSSSRKAIIRSIRDVSAGSENNVLSQIYAYDKLRQISNNVLELA
ITWCREOYROMVMWYEAIAKINPTVMTAIGKPVSRKALGDVISTECINVDQSSVI
HKSLETNNDICYSRPVTFKYNSSQLFKGQLGARNELLSESVENCHQNAEFTFT
AKNETYHFNVHVETPLNNISTLDLTALNLTFTENIDFRAVELSYSSGERKANVF
DLETMFREYNYAQISGLRDFDNSQRNDRRIIODFSEILADIGLSQKIVNVVASG
AFSLFGIVTYGILNFKNPLGGMFTLLIGAVIILVILLVRRTNNSQAPIRMYPDV
EKSSTVTPMEPETIKQILLGMHNMQYAYKKKEQRAARPSIFRQAETFLKRKSGY
KQISTEDKIV"
20507..23536
/note="Orf 09; KCRF2"
/codon_start=1
/product="DNA polymerase"
/db_xref="PID:g60330"
/db_xref="SWISS-PROT:P24907"
translation="MDFINPYLSKRKPTDTKPKLHTTQSIQRLVVKPCFRNTEKGV
SVSFSALPTYFFKGNEKVIYLENGKMMHLRRPCKNALLEQSIYTHIYDIYETYS
KQALNDIPFKQTDIIPNGTVLKLGLLEGASVCFVQNRNYFVYKVPDEGNNITLYI
KOALNEKSPCAYOTEAVKKILSKRYDEEHDFKVTVSSSLSVYKISDSLVNGCE
VFETNVDAIRREVIDNDPFCWYTKSACPRITNRDSDHTDIEFDCGYDLEFADRT
WEPYNNIMSFIECIEGKGFPCAKNEGDLIIQISCVFHAGALDTRMLLSIGTCSA
VENTVEYFEPSEIDMLHGFSLIRDFNVEITGYNISNFDPLYLIDRATQIYNKLSD
YSRVKIGEDYSYKIDIMPLFMSPGPRAKIGLYCIDSVLVNMLKLFMIHVEISEIA
KLAKIPTRRLVDGQOIRVFSCLLAAAEVNLVPSNDVNADGFGQATVNPIDGY
NNAVLVDPSLYPSIIQAHNLCTYLIPHALHNYPHLKSSDIETFMSSGPIHFVK
KHQASLSRLTVMLSKRAIKRLAECEDLTIIIDKQOLAIVKTCNAVIGTGV
ASGLLPCISIAETVTLOQRTMLEKSIFFIEMTPDTLQEIYVPHIVKPEDAKFRYIG
TDSLFCVGVSVDTVYVFGDFAAFTSEKLFNAPIKLESEKTEQCILLAKKRYIG
ILNSDKPGLVDLVKRTACKVQVTSKILNLKIDPEVKAQAOLLCTKDPDAPFRE
GLPDGLFVADILNESHKRLRTGOVPELFEFSTELSRPISSTYENLPHLYVYKII
TRHEPPQVDRIPYVFGVTKTSCISNMAEDTYTVQNNIPIAVDLIDFKLLHGVANI
IQCLPKDSKSVLYNVEFVPLVPLFSEYELLTDHSVKA"
23573..24796
/note="ORF 10; KCRF3"
/codon_start=1
/db_xref="PID:g60331"
/db_xref="SWISS-PROT:P24913"
translation="MEPSISFVGRWTVTVNSGLFSISNHEEIPATPFARLVHLPH
SIESIDIRISHLGVLTLQLOQYNTNSPAVFSYDREQACKPYVLPVPSDFINPLI
VYVKGPIILIRKNNELKHFVITSPRADISFIPPEDDQTSKSLTLAGFQKTEINEI
NICGPTMTKEAGSVMLFCRCKIPFPYNTSFDAKDSGLILETLIQEAGDYVMTV
HLCSPKPAHIEVQVHLVNLSPLELVEFKHALPWEAPSGSVLVIIEADKLIKPGN
SVEICISFVNRGLVSDQPALFVASSNHTYVVKPQIWPITPLSITVYVNSRRII
FIKRGFVAVAVPCFFHLKAPQDCCEERVILDRENSSIHWSVLIKPGAGGPIVHVH
VALKEINLTETPEPNE"
24799..26016
/note="ORF 11; KCRF4; similarity to Lf2 of EBV Raj1
insertion"
/codon_start=1

/db_xref="PID:g60332"
 /db_xref="SWISS-PROT:P24914"
 /translation="METPKHIRGVWRDSVLKHKREKFNYSFWNCVSHEHFISITNNR
 ETHLPDLDLLCPRAILFLKQPSFKSGSRGPGCTVTSLVYVGHSGKNLIKTKP
 MVSCEDELIFKIFALACTIPGSMELFIPFLKDLGLYLCLLEDVTSRIMSTS
 CQMGTYYLASETPQVFLKGGPVNLKHEPMPYLMQAOKFPDFDKMARVHTVQNEVCVN
 SIYRGHNHVKAIQKDSIEDINFLDTVVVGLTNTNRALVAFENPNYFSCPDWKKRQIP
 IYDGPCRIPAPRALPVKNNTYYSIYPNATAIITNSDCADFHSDEWPKGTAC
 IYVNTSIVTSIGSTHGEAVFILAPKFFCKIMSKTHVQKLSAICLPGNVTINS
 NKVPKLADLSTVK"

CDS

complement(26045..26554)
 /note="ORF 12; KCLF1"

/codon_start=1
 /db_xref="PID:g60333"
 /db_xref="SWISS-PROT:P24915"
 /translation="MSSIQKCLICNIGEEELLQACDCPSVHHTCLOSHQCFKSS
 HCTFEKRYKYNWNCSSKSCPSVLQANWVLCVCTILCICLIDILITIRLW
 QSSVUCYEYINTFYFLVLCGTFSAFYLAAYDIFEFPHSLCSFIWNLLKISQSPCE
 ASKNALKIL"

CDS

complement(26936..27391)
 /note="ORF 13; KCLF2"

/codon_start=1
 /db_xref="PID:g60334"
 /db_xref="SWISS-PROT:P24916"
 /translation="MTFRMTSLVLLILLSIDCIVKSEITSQTPRCCLAANNFPFRSV
 VTLNRNNTSSKRADYVNRSTPWTLHRNEDQDRYPSVIWEAKCRYLGCYNADGNV
 DYHNSVPIQOEILVYRKGHPCPNSEFRLEKMLVYVGCCTVPIVHND"
 complement(27674..28423)

gene

/gene="IE-G"
 complement(27674..28423)
 /gene="IE-G"
 /note="ORF 14"
 /codon_start=1
 /evidence=experimental
 /db_xref="PID:g60335"

CDS

/db_xref="SWISS-PROT:Q00997"
 /translation="MALDLRLNKLHTANFSFRIMIWMCLALPTDSKPISTTEAPIL
 NITQPSLNISSEPLEPKNCTTFLDLWQRLGENASIKDLMLTQREVEHGRM
 TTLPSRPSKVEEQLOLORNLPTANGPPHYKLYRLNRLNEAPKADYNGKIQDF
 DPPLPYTGANDDGLVMYNINGKHVRFDSLSTWERIKRISGTPWCINKTSEKAAILQKL
 LKAEEKRTTSKRITELEEQIKELEKTSTSP"
 29231..29596

CDS

/note="ORF 15; CD59 homologue"
 /codon_start=1
 /db_xref="PID:g60336"
 /db_xref="SWISS-PROT:Q00996"
 /translation="MYILFTLVTFCKPIHSLQCYNSHSTMQCTTSTCSTSLDS
 CLIAKAGSGVYRCWKDDCSPKRISNQLSETQLKYHCKKLCNVNKGKGIENIKRTIS
 DKALLLALFLVTANNEPL"

CDS

30322..30804
 /note="ORF 16"
 /codon_start=1
 /db_xref="PID:g60337"
 /db_xref="SWISS-PROT:Q01001"
 /translation="MSSIKFQNIENILKCCODNRHSDSVVRVAVSHVHOYNKEPAL
 MPDESICVHDRKKTGEAILLTETHTWKGKVMLSFAVLQIDIEYKCVASML
 SSYSRSVGANWFENGKESLVEFCNSIMPQNPFLVLPFAVLGLVLMQTLIK"

CDS

complement(30837..32264)
 /note="ORF 17; similarity to other herpesviruses (HV)"
 /codon_start=1
 /db_xref="PID:g60339"
 /db_xref="SWISS-PROT:Q01002"
 /translation="MSIVYAGFVDVAVPKVDPVLYLNLDDVSKCLPTKPIPLNIE
 HLPESTIGHTGLVATGVGCVGHSEKFLHTENLFSNCSVAQATSKELPYOPLL
 EMLHTWLPALSLVACTQAQNAANTMFQHSVLCALGRBRGVAVYSMNLDAISQFC
 SISOAEVNIQDSKNVDINSLPKPVFNIDPHILMAKADAGIKDRLOLLKTDKGA
 KIKKTYLKASIGAPVTEISEDNMQGVIPQSGDDLLISVPKSTFSLMNLNDNFK
 QHRPACFPQYFSGOAGAMPYELYPQYSGNIGYMLPSGSIVFAMFPSPRNKKRRE
 DFDCCVFGESLYKDVNLTKNISLODLDLKDLOAAINQPNRYPPHFSNPYSIDLP
 CHASYFRVAPYAPKPDQHLLOPLACVQQAQVQPNYAPPPTEGASNEAPKFSVQEPV
 HIDASFAODPVSKLQKMFCELLNK"

CDS

32257..33027

CDS

/note="ORF 18; similarity ((sim.) to other HV"
 /codon_start=1
 /db_xref="PID:g60338"

CDS

/db_xref="SWISS-PROT:Q01003"
 /translation="MDILGRFVRESKMSPLKATIMEKVLRGOSLNSFHSEERLFLHL
 VICKMYDFCLNVIKLESIINTGRDNEVSRKVPVEIWKIDACKSIGVENTMLJD
 DSRGQWLHLNSNLDLQMSQFIFSKLGIKHFVAKISQNIITDGNLFNLGSLVPYR
 LILLOFCILFWGKEQEAWRFTGKIEMLYLITLGHLLIQTLFIQAASTGVCGLP
 EIIGDRLSYGLGHTYNTDLOHIPSLLDLFIENNEY"

CDS

complement(33024..34655)
 /note="ORF 19; sim. to other HV"
 /codon_start=1
 /db_xref="PID:g60340"

CDS

/db_xref="SWISS-PROT:Q01004"
 /translation="MMKLEKVIILRONPSVFLNGTAFWTPHPONILHIDRNSLRETKK
 NASLYTRLNLNLETEQIKAMINVELDKMDQHVKRSSIITRDLIEIENVEKFDSP
 QLPSFPPKLPSTQOPTNKTNYITVAPGDPGFTVESKFIELVSLYTNQOQM
 LPSYGWPSLTDIAMQRRVFPKELRGTNYQNSTSLKMHAVLTITSSATDGFYSV
 RHISDTSALVILNVAFLKTSAPIPVYVEELNLEAKLGMFVDELKNTGGGRSF
 SPOVNEATSIAPNKDTKYSOHSFKSIYSLLEAGLLSTKSHETNPKTVDLYTIT
 TEIFGEDIPMASFQWNLRVGIIAIEFVITYLLETSSQISIHSTRRLNLSLLGSK
 FKASSTGLLNQIVIKKGQVFSFLKNYIVPTLTHKNKVPSTFLPGVTLLALESLAT
 AVDKPFINLGNRFQDIFEINQKFTFKDPVSLMAAQATLRFKVEHGLSNILNLSPT
 TEATEIRIQFGGEDDYTLFYVLGCLPIAAV"

CDS

complement(34285..35196)
 /note="ORF 20; sim. to other HV"
 /codon_start=1
 /db_xref="PID:g60342"

gene

/db_xref="SWISS-PROT:Q01005"
 /translation="MLSVIKORDEKVLHLPNKRKIAGNAKHALETYKKLAKYTVSASI
 EKFLSISHPCLPRAKTLFEVSLGNRIADCVMLTSGETRICYVLELTKMFSNLDL
 ISDIRKSQSGQLQADVNFHNYAPLGRQAWTVLPILIFKSQTKLTLHETPKF
 PVLNLTSEELSCFLWSRADVETIRKKIHUAPKPKRIFKWDSDLDSTDEHSAYROKL
 IERNKKCTFLQNTSKFRDRTNKNNDQLRQANARPCCKKQHNKRLNRKHHGG
 KYSRLTJTTSFSEAAFSNYPVSTHKL"

CDS

35195..36778
 /gene="TK"
 35195..36778
 /gene="TK"
 /note="ORF 21"
 /codon_start=1
 /product="thymidine kinase"
 /db_xref="PID:g60341"

CDS

/db_xref="SWISS-PROT:P21293"
 /translation="MTGRQPPKKNNDIYDPKQPKNGSYNDYPTSTKTRNNKQ
 RKDSNYPPTRETFEPDLAADPIYSVPPSRVPHKLVKNVKNLPITASNSVSELL
 SLHDETQVTECEAPLIAKSPDITVYKMFVRPKHTLTLEKQKMFTRKKGSVFK
 IGSNMLEFESLKSCLHNSKSPDEPDGLVHVPHLLPPKQDPVPAFFLESGI
 GVGKTLKSMNGILGKGVLAPEFIATWVFSNSLEEVYKLTLPARKVGRTSNAK
 LLACQLKFAASPALLAKATATDRJSSPKNSLLSDMWYFDRHPLSATVFPYMHFQNGF
 LSFSLHQLWSNFKASRGDNIILLNLSQENLARKVRKRNKEKSVSIEHRLNNGY
 HAVYCAWLLVQNFTEPEIVEVCNAKHITDLSKSPSLAKHVSIEDMLKSSIFNTWI
 EMTKAHRDSTLMCELLTFCKEKLVQLIHVNVSPPTDDIPGLWASITYTSIRNSAIK
 PNRYWALDELARTENSO"

CDS

36775..38928
 /note="ORF 22"
 /codon_start=1
 /product="glycoprotein H"
 /db_xref="PID:g60343"

CDS

/db_xref="SWISS-PROT:PI6492"
 /translation="MTILQFLVLEALCDYQLPKPRINKPPAEERLKLNGYNT
 TLIEFDGQVSFNLMWTKIEHIPHDELIELREANVTEPLVNTLKRSDTYRPEYNV
 HYPEGNSACALPYMSYTIQDQWNTTGLNGFGIPSKTVLNEFFYQYVYTNRO
 FTYETVNLCLIGATPAVPTISCHITNPLVSVFEFKEDSLTLFGHSHVLPK
 GHVYNDISGASNDVSLVISTYDLFGKHVESFKEDIAKVFREIETPLPTIKNLQ
 DEMFTLEIRDCGNINNVKPTLFAKAVAHFLVDSLTQQHILLNCFANYMSEL
 EFLRKMESCDEFEFDPFYTETLAASQALNVKPVITISLHQDILNMLSLFSLR
 LKHVSVAISEIIDLISHIYATSYTMTLSERKMLDAYIVLNDIMHKNLTPKQ
 DLLPYVLSSMCTSLIEIGNLLHFGOKDLDVYTFSPCYLSLRFDFTKEKLTPEPO
 SLLIAQKENLNTNGFPOTLHRHHTSLEILPIKICKLSLTDIILSLKNTYVIS
 TPVNSKIDYSEVFLKSMISAVNNDCKPYOGGSAHQPVIIVNTVTPRGGCPYC
 SVPLSYDESQGGFMVITDVTYQENLFTESPFGDGNLHIHILMLNNGTVIEVR
 GAYRARLVNFIIVMVFILVGLVLYLFLVLT"

CDS	complement(38921..39682) /note="ORF 23; sim. to EBV BTRF1" /codon_start=1 /db_xref="PID:g60344" /db_xref="SWISS-PROT:Q01006" /translation="MYTQRSSSTKKVYKINEANNNAVVEKKNPSSIKRAELKTVLI PKPIIKGIIICGLEKAYIAKAKHPVLVIGKCPSPRYHTINWTKSNALILTPSD INKTOLLHLLSLDYLGLKSNQAQFALYGVLEPMPTSKEMLSQDLFFETKRTED IIFVLNTICHRFVHPVRASVTQVFKDIVEKVFELFPKADKSNINFAASVVEIICN GEPFSKVIQVNAVMDIKETWNNLIKIYALLT" complement(39692..41887) /note="ORF 24; sim. to other HV" /codon_start=1 /db_xref="PID:g60345" /db_xref="SWISS-PROT:Q01007" /translation="MSMFLLPVYNIPPAKGLKVKVQTRDCNFNIQLIFCCLATNQVY SSLDNQLDCIMGHSTFYTCRAVRULLGSMNYPFIDTLNESATGTRGPMYNGPLI INTDSTYKLTINCKNRPVYSLETTDPHPPLAYRAIYFPDLEQTPIDYCMFKI ICRYVMSLEECYEFYLATVSPFVNTCKNKLRLVSAKLTLPSTVLAIPNPDDQLE FFKYSILSFOEWSLNSLLTKKKIAAVHSHPHIVIKLCSONAFKEIKITDANFE MOQTVNVMPEHAETIKRPGSRPLKSVLLPDGVKVIYPPSLPIRYVTMCLASVA AYSDTISDKQDIRTASALYSIFERKINAPKOKKKDNVNSKVALDIFRNYIAQKT EDEFMTYRPHRLGNIKNVNFNTNMINTKIFTHSVPTWYKSLMDIPRLTKNFEY KYSVKEPSTVSYFCENMGAAININISGDLTLFHAMGNMKCYMPKNTLIPVSL SNWNTLDLHGLENSQIVTRGDRFVWTFNPPSAVTKLGFNVSWFKAATATISKYGL TSLTSHVTEVSPYIGNNSARLSLLKNLPSVLESRRNSOIQTLHKRMLCCLVALCSF LRSDSTVRRLAOKGFDESKTISHAKNHKDCALLGYKCNMIPKVLNSKTRLDE HGRNAPLAFISATGTFVHPKIRAKLNLKLVRLTGLHWRRTYKHYVNTAQ" 41893..46008 /note="ORF 25" /codon_start=1 /product="major capsid protein" /db_xref="PID:g60346" /db_xref="SWISS-PROT:Q00999" /translation="MEVENPYPMVSDANLLOQIKESSAEGLFKFSLLLGKDVRES GVKFEALLGYVNTATVKEFLETS LAVSCVNTFEFLKRTMDGKIQKFINVPTIAHD GRPQKQKQPIIMKATNKHIGAEIELSTODLEFLKSKETPLDVTVEYGVAKTITSA LQFIDALERGLDITVLKLRHAPLFLIKTLADPTYTGERLKNVKSOLISWFKTH LVNNSFLDKSEHLPHSRQVYVGLILEMIGACVKEVFKGISYISTANGQPIISVLET TKVMKLVNIGQDANSIMPAAYANVVRGENLVTAISYIGKAMRNFHFSKLVDN PYSNLDNDVDFESTGSIQKTPISVVMVGNKLTALESLOQRYNETOLPYPLNRMS HYTYPPGLPSPSPISVMSVGTENVLHQSVAEWIVNKNNTLQCFNTQNALKSTIC HPRMNSPILCARALGEAFPDVHNLVYIGRSEDAHTMNLQIVYVDNKHVAHVSL AOKSMTHHEVLHPTHEILRTEVHPFEDVYAEHOGAAVOYRATHNLSGNLPPPLA PYSFOCRQYQFEVASGLNHVIDSTWELIQTAFDPAYELLCYTESMHHGQEEFV MNPILALCIQTWNNSGRUAFINSFIMLAFICTHMGNGHISKDAYSCYRIYGELIA IQSLYRLAGHENAVENIGQLNAILDKDLPLFPAYNDIFTNLLKSSRPVYKIGM EYDDNDQNCINI REKMDLVGNMNIYQQRNNTDHSRYVLDVGELOENTYNSVL EKIFYVLLPVCTNHVCGMGVDVENVALTLTYNGPVFASAVNQDADILDHLENGTLR DVLVASEIRPTVGMIRLCTSFCTCPILTOAARIKTDRPGQNIVHTDQYVHQIVL VNGFAAPATADSRDAHLFYVPVKNLYCDPMAATLHPVIAEFLITEIPSQRNAVY FNLPPRLIAEYBHWKSPSSVYSTCSQPLSLTAMHLKSPVFSFICQSRKHHP GALTAVRTDEVVAEHIMTSSAKTSVF IGQPTVHRKEVSDAVFDDINHELASLDTA LGYSTIYPAHAAAITDMGIHQCDLFAMPEPSSAYNQOLNEYIKOKIGSDRVYGMPL RDPREYMGKGNRVTLPLGSHQOLATCEVIMPTADITYPOSSNSPRGRASCYVSDCA YNNESAELYHLSLDPDCYEEFSTINPHASQIGSLGDVFEFNSQHRQMAGPTLYSPCK QFENKAILKNKFLTYLVEIVNRLTGAPATNTDIFYVYVINGDTVFLDQPCQFLOE A1PTLSASHRALLDEYMSHKTTHAPVHVNQYLVVEVAPMKRLLKLVGNKTY" 46024..46938 /note="ORF 26; sim. to other HV" /codon_start=1 /db_xref="PID:g60347" /db_xref="SWISS-PROT:Q01008" /translation="MLTDKTIIVLSRSLFADEITLKQKKIGSLPLQDPHKLQSLDT LGLNAVSRDVPDYPVHMF SYLSKCTLAILEVNPDLNLTLPDQSETYQIKNVPBM FOWDGSNTVTPVFGROQATVTLESNGFDLVFPVSPVSLAQAIIGKLLNLYSR LVESDPEINIEVNMVYTNTHGRHYVLDINNHNNEALKSLDLDVATYCLISALIP RACLRVLTLMRHDQHELDVFGVIVPREYIEIDANALSIGDDITRTMTFTITYLQSL SIFNLGAKLHLSYSYASETQTACWISYC" 46945..47787 /note="ORF 27; proSite motif for cytosine-specific methylases" /codon_start=1	CDS	/db_xref="PID:g60348" /db_xref="SWISS-PROT:Q00998" /translation="MQEDIPNLYVSREEDGRIVEMIVTEKHKYVYYPYPSNKSFTT IEDTKLNIANFSLIIOWIYITLYSQAIIIFILYCWALNPIYRPGNSVLGGLGORVP RTVWNPHNIFEGCDKSIICKRLPMPFINNTHGKIYPNFTKGSANTYKLAALERL VGLMNOQCNVEIISOKKTSFSONQVTFEENKSDAILALILVLOKGNCHPESVIVKSK IKLSENYGVNYRQHPQYFTVCNASWADVINNTNFKTSDRPCNTNLLRYPOHVPII INVK" 47862..48143 /note="ORF 28" /codon_start=1 /db_xref="PID:g60349" /db_xref="SWISS-PROT:Q01009" /translation="MATTVPSSSLWTPSMKTSAFGIILLICLLIIGIYKIVIRICI IPAAWTAPGELLVHINYKAKIYYNQINIFYVLLSGLKMLCFEAVAK" complement(48089..49231,52315..53223)) /note="ORF 29; conserved spliced gene" /codon_start=1 /db_xref="PID:g60355" /db_xref="SWISS-PROT:Q01020" /translation="MLLLKAKAIENLSEVSTQAEIWDMDMSTPTIITNSKERT YSKTVIPSNLYSSTISFCKLYKPLTLNQTPQTGLRLPHEKPLILQDLNLYVK LLTSQVCHDTEANTYNAAYQTKTSMECTYLELQFVNLNLSFNGCVKRSHTI EPFQOLILHTFYFLSIKSPESNRLFDIFKEVGLREMDPDLQIFKQKASIFLIP RRHGKTIWVAIISMLTSENHIVGVAKQHVANSVFELINTLOKWFPSRYDIK KENGTYIKSPDKPFLMCACTCNKNSIRGQTNLLYIDEANFIKDSLPAILGFM QKDALIFISSVNSGDRATSFPLNKNASERMLNIVYICPDHDDFSLQDLSLSCPC KYLIPYITIDETIKNTNLDGAFTELMGDMGSIKSNMHSVISEMAITQFDLC RADTTKEITQCLNSTWYIYDPAYTNNSAGTIGAILTFKNSKCIIVGMEHYF LKDLTGATYQIASCACSLRASLVLPHQCHVAVEGNSQDQSAVAISLINCESP IKVYFIHKDITTMQPIYMLGAESIAFESEFYALNSGTISASQSIISNTIKLSPD PISVLIEQIRISIRCYPLRDGSHYCAKRIYVSDVVLVAVNAYFAFATSNKHIFKPLNS T" complement(48089..49231) /number=2 49249..49476 /note="ORF 30; sim. to other HV" /codon_start=1 /db_xref="PID:g60350" /db_xref="SWISS-PROT:Q01010" /translation="MSKPCISEDFKDCQAFKAPLPLVAEISKSLANIDFVDTASQ QVQLGFLDLIGTCEKFCVLCPTNSANLN" 49443..50069 /note="ORF 31; sim. to other HV" /codon_start=1 /db_xref="PID:g60351" /db_xref="SWISS-PROT:Q01011" /translation="MSTHQCFKLNIVVQYHKVFLYQCCCKKHYHICDGSKTCIL INTGSLVCLTGNCFINNFQENTFLQADNIKESKQFDTFLFANILESUKQDLNIF IKSYGLEEVSKAILDNNSLKPICDLIRCTFHCCCHIFGIQYGIICSMYHIITS IYSTRTVYGLLFCKTKKKYKYLKRMRETWNSTLITGDSQCQDIKS" 50015..51340 /note="ORF 32; sim. to other HV" /codon_start=1 /db_xref="PID:g60352" /db_xref="SWISS-PROT:Q01021" /translation="MDVHFDNRFLSRPDKVIVHLLLPQTFFESHHPFDPNLTFSWO TRFHGLDPPTRVYKIVGKLYFSLPGPSKACIYAVSLVINIENNYNPMDFMVLKI LLNENYIYKFFHMFELFPYIAPNSLODTSIEEPKKNVDHIDALINTSPNTLYKNP LGVANLLQSRPSSPRRHFALEDPGKVRGIEQPSLQTDKAMPKVSVKHKWILNS EPVTCVKHIFTKRSYFCISYPMNSQCTSVLDTISSBELAHVSPTSVLDAAEKAFLF KHQHRFVNTLHVCKSNYAIEQHPVPILOKDEETASIKDHFETCFVLESTVEAS AWYRAFARYLKNKPAFIWYIRLWEEGVHTLKGSPDELKEDVCEEKWHLLSUNKEF IHGIRNQCQGVLDVDSNLNWLILPGGFVYKGVHNTTPEDILFVGARY" 51333..52325 /note="ORF 33; sim. to other HV" /codon_start=1 /db_xref="PID:g60353" /db_xref="SWISS-PROT:Q01022" /translation="MDDFRNKLNFNLNECLWNKNVACTSFTKYVCATTAVSPFKPI SPQGVDPDKHYNVTLILPKKSHPIYTIINDLAVDCCSTELQVAPVPCSPSLIY FGLIAPPHNVNVOIPANLSKSKSHLTKNQVIFTSKVTHPERLPDGYKSLAGACA WYSEGAIFQHFLSTDSLNSLCPAFKFPPLSRILSLTLTRCDDLSCVPCYGEKIHVNCQS	CDS	complement(48089..49231) /number=2 49249..49476 /note="ORF 30; sim. to other HV" /codon_start=1 /db_xref="PID:g60350" /db_xref="SWISS-PROT:Q01010" /translation="MSKPCISEDFKDCQAFKAPLPLVAEISKSLANIDFVDTASQ QVQLGFLDLIGTCEKFCVLCPTNSANLN" 49443..50069 /note="ORF 31; sim. to other HV" /codon_start=1 /db_xref="PID:g60351" /db_xref="SWISS-PROT:Q01011" /translation="MSTHQCFKLNIVVQYHKVFLYQCCCKKHYHICDGSKTCIL INTGSLVCLTGNCFINNFQENTFLQADNIKESKQFDTFLFANILESUKQDLNIF IKSYGLEEVSKAILDNNSLKPICDLIRCTFHCCCHIFGIQYGIICSMYHIITS IYSTRTVYGLLFCKTKKKYKYLKRMRETWNSTLITGDSQCQDIKS" 50015..51340 /note="ORF 32; sim. to other HV" /codon_start=1 /db_xref="PID:g60352" /db_xref="SWISS-PROT:Q01021" /translation="MDVHFDNRFLSRPDKVIVHLLLPQTFFESHHPFDPNLTFSWO TRFHGLDPPTRVYKIVGKLYFSLPGPSKACIYAVSLVINIENNYNPMDFMVLKI LLNENYIYKFFHMFELFPYIAPNSLODTSIEEPKKNVDHIDALINTSPNTLYKNP LGVANLLQSRPSSPRRHFALEDPGKVRGIEQPSLQTDKAMPKVSVKHKWILNS EPVTCVKHIFTKRSYFCISYPMNSQCTSVLDTISSBELAHVSPTSVLDAAEKAFLF KHQHRFVNTLHVCKSNYAIEQHPVPILOKDEETASIKDHFETCFVLESTVEAS AWYRAFARYLKNKPAFIWYIRLWEEGVHTLKGSPDELKEDVCEEKWHLLSUNKEF IHGIRNQCQGVLDVDSNLNWLILPGGFVYKGVHNTTPEDILFVGARY" 51333..52325 /note="ORF 33; sim. to other HV" /codon_start=1 /db_xref="PID:g60353" /db_xref="SWISS-PROT:Q01022" /translation="MDDFRNKLNFNLNECLWNKNVACTSFTKYVCATTAVSPFKPI SPQGVDPDKHYNVTLILPKKSHPIYTIINDLAVDCCSTELQVAPVPCSPSLIY FGLIAPPHNVNVOIPANLSKSKSHLTKNQVIFTSKVTHPERLPDGYKSLAGACA WYSEGAIFQHFLSTDSLNSLCPAFKFPPLSRILSLTLTRCDDLSCVPCYGEKIHVNCQS
-----	--	-----	---	-----	--


```

/translation="MEDIEGGISSDDDFDSSSSDEESDTSQIMKSDVMTASPP
STPEPDSVASTNKRERQSPITWEHOSPLSRVSRVSPSRFGRPRISSTSGR
SCTSWADRVREAAARSPRFRKPYSHRNGPLRNGPPRAPPLKLLFISILPKSG
EKPLFVPSLPCQAEKTKNDKVLAMQAMHDVPLSSKOLTANLPLVREKPLLSIV
RTPNYYVWVMKRETIASANTVAALFDESICWGQOYLKNDFISENGKDIILYS
SALLSOLVHKMKPCHCLMOTPDQHIQVQCYLCSNNRDLDAVYLOTYIKSP
IVLLAYAVLPAIICITKNETQILYSHCMRILREYRPGDVNMLIHESLQHLNCPSS
TCAYTTRAVGTRANTTGLFPLTQ"
exon
<78291..78311
/number=1
exon
78395..>79627
/number=2
CDS
complement(79986..81059)
/notes="ORF 58; BELF5; sim. to EBV BMRF2"
/codon_start=1
/db_xref="PID:g60379"
/db_xref="SWISS-PROT:Q01053"
/translation="MQEMQTQWLNIRASCLGALATPTFIKSLFTFSIFSL
QISFVWALGAHVTILCYCIITFKEMSYIIEALGACLFNMLTFEISHLWAPLY
TLPFVFLGICLSLWPIYIDIVILCPETIKYIELGFLNMLYVVMANRWLYSV
IPMCPVLGLMGVEALKNEHPVENILITCKPIFTAKNKYKTKTEVMKLVNIN
LGAVLLAVFEASVSVQRLDIFVGMQNYLFLFVSMCCCMFSLPSNAICVVLST
FVCTVGNIVSVALTCCNKSEQL"
complement(81056..82162)
/notes="ORF 59; BELF4; sim. to EBV BMRF1"
/codon_start=1
/db_xref="PID:g60380"
/db_xref="SWISS-PROT:Q01052"
/translation="MADTOTHSYGVGFKPCVLOEHTKIYNHIKAKVKGIIQIEGL
TSSPSLRWSSIGEGILSFKNVSEFVNCNMPENVVSFRNISPGGNTFLYREL
FGCNVATLAFMSNRSGKPEFTKAKLEYCDKYSTHRTSTIIPAGLFFVFNVNSTV
LSRVLSPTAKMLQWLRHKHAKHTVSMVNTLAVLITVEDSTKIDVLIHVDMS
LCNNKSDKGVDQVQTKNITLALSAALGCKIPGVFCIKFYESEVELEVCGTPYK
LGSWDTSLQVTLFCPLNAEDTQAQVSDYFIPSPSACLADPADTQEEPTLSCNCD
SVSENPALPLKRRKREKDKKSKIKIAFNPLI"
complement(82273..83190)
/notes="ORF 60; BELF3"
/codon_start=1
/product="ribonucleotide reductase, small subunit"
/db_xref="PID:g60381"
/db_xref="SWISS-PROT:Q01038"
/translation="MSVKKYLTCBHVGFLELTKETWKNRWPSPVPLQDVGCCVDM
LNERDEFKFLFTFGMAEKLNNIEDLSDFSDHSDISHYAEOMAMENIHGKVA
NLNMLFKNNITVSYACDMNDSALQEKRLWLNREASDAEKILFLFVGEIGF
FISFSGIGLFRVGMGICLANDIARDMLHTSAALLYNTMTKSSRPSDWTY
KLFRVAVEFKFAAGYGVSLVNVNHEIRQFLQATADRILESINLNPYISUPPENC
PLAYTSKTSVNEFERDNDYCTGLNDL"
complement(83196..85499)
/notes="ORF 61; BELF2"
/codon_start=1
/product="ribonucleotide reductase, large subunit"
/db_xref="PID:g60382"
/db_xref="SWISS-PROT:Q01037"
/translation="MSQETIISNLIDMLKVSAGWDRANEISGRLFHKLMDMSTETPI
SQYMSFLPLEPHLEIFTEQYQDEIEVLEYRASDFNCLNGLLPKAPRYDTV
LPRTBMNGKISPIHFARIAAYCAWNCIMCEPLKDTLYVQVQRDNWNPILKIDQIF
KYFYKVISQLVCCATPMRSAGVAGENSLSCTIAPTLOTETKSTISSIFGELAPLLA
SRSGVGDVVTKFSFGKNIHSCKLINAQVEFFNDKSVRPVSVATYIEVHCQIHEFL
SAKLPNRPNCSNIFQGVCPVSLFFKMYESDNGMLTFDQDAPNLTLYGLEFEE
YLRVSEKTKQSVLMSFLINTIKTGSPYVTSKSNKHHWTYETQGEAINCEN
LCAEIVQPKQFTSTCLNANVCLPKCLNSNFPYTSKTAQFDFSKLEYAVQAAVFI
NACILSPSTTSATVQGRSMRGICGLADVFSEMGYGLDLESECLDRIETMYI
TAVKTSSEICSVCKGQPFAGFRGSKLAHGVFWATWDAMPQVRPMQWIIHQDNKKE
GVNSFOIALMPTAGTSQLTGYSDFPYFPANMSSKVSNEEMKPNITELKNVKKPF
LCTVRYIGDVSMPDVSRYRKHETAFDPCPEAQRRASIRAPYVDQOSQSLFTI
EENVOAKYLKDLGLFGLTKITMYICRVKTKTKLLQLECLKLDHETKKAQIVLA
DIARELPDSHKTEDACPLDQSECAQ"
complement(85499..86491)
/notes="ORF 62; BELF1; sim. to other HV"
/codon_start=1
/db_xref="PID:g60383"
/db_xref="SWISS-PROT:Q01051"

```

```

/translation="MPKVNENSFKKIEKDKVLKLLPPRPKHIEISEGQFAKSIROALS
KYVTSVNHPIDEFSSQMDVPFROPLVADFLIOAKT IHQKEP IGSFLTFEKOEDSGSS
IDILFSPVSEFKISGLDAEFAPSHRISFVYAGNCDNCEITVPNLAELLEGKHFQFL
TPVGLPVENTISTFVTRIKSVVKGVLNKKSPPENKULLPSDFIDUDEPNIDTLAP
KHLISFYACVTYPLVHNVPYMLTFTKSNKSIDRDLFCHLKYVSDIIRTYIAMOND
FYINMFGKALCKLSSSSETVPRRGNYHFRGSSLTVEVPDFVADPGPMWTL"
86498..89197
/notes="ORF 63; EERF1; sim. to other HV"
/codon_start=1
/db_xref="PID:g60384"
/db_xref="SWISS-PROT:Q01055"
/translation="MSVPLTELTAKLLEAASNELLTKVKMDMELNYLSIEQLSAQN
ITDFNLILKQTSOYSTFIHOHEFLYLLKSLSTFSTLNYDLESIKOYMNILNVCDAV
TIHATSSNENLNQTVINHIIKQYITSNATFTGLSEPIVNVNISTFRCVEVHVCFQ
CIWHFFQAOQIPQINGALEKWLITQHFLNLDITAFNSLKDQATYILTHEKHLLFVP
LSSEYSLTDLAKNALNIYTSFTNTITKSNVPVAFSAKELTDATPELFFYDFI
IEALYHHSYVNPQNIIEQFISKNTQFMDLTCNTIQKCSNKSLTSSIRIKELLES
CGLTECHRLQTSVLISNVSTNSKNGYETFISLSQVLSDFYKCLFYFRGPTS
IGHSKITELNVTSAISETLSHANKSWKLANMSPFLPKAPSKILLETHTHISPYL
KMSAFSIWAKTWNITWLDATSPHTQHTFNHAPVISQSEVQKICEMLQGLTBYDS
RIVNSHLFAEEFITHIITLPTALONKQKRALFQRLWLIVFASDAKGLYRIRP
LGLYFOGIEIFHDSNAAAAILNVDLYNEIQLOIYVPTTPIKFIQELFSIKY
KPOSTELSKIOKFAETECVDTLPFIQLGTNNCNTYVHIENYINIQOGSPAR
LDTKALHAKAIQGLKESWTTISQSKELQATYIOLATILEIEKISQSIAKVS
NPNEIKANNFLQCFKKNYNTANLITSHSENLRIFYRQIFPELIPITITVOKILNF
TETDDPOFPLDSQPLYSHTNAPKSELSTDFNRLLEFANPFEFAPSSIKLHYSD
TFNTPQVNNKTYHTTYIADSPAELQFTHLTSAILDALS"
89200..96609
/notes="ORF 64; EERF2; sim. to other HV"
/codon_start=1
/db_xref="PID:g60385"
/db_xref="SWISS-PROT:Q01056"
/translation="MDHPLFKLNLEGIASTHOADEKYGQYAGSQCLSNCFVLSVS
YQDTEPMVSLHGLNEILYKAGIDFILLRRSGQLHNOYAOHLHPIGVYAGPKWACFI
VYNTSEMGHGHESPINEPFFVASLKLKNYNTVOYFLAICHKSGMLIKOKKIF
LFDPHSCLPVNSPAHFVSTSNVDAEYLSPPNVQYIGSLYFVPRKYGHSHTMIN
HYRVNIEKHGPNIDLTQEGLEIEISPPNPKPTSTQPKPTPTPKPATPKAPT
PRKPKTESTIPYDKSKPKPKPKTKSKKVKLTQKALTQPKHTIEHRELAPPI
TETVEDNTFNHPVERTPGDLSGINSITKREDDLEDDNDYTKLEDDDDWDID
IPIPELDTTETHSDQETIYMGIDENHDWSDDDIDDTLIDISFIDNLNITSIDNI
PKNTPRIDKTSNQPIKEKALHSIDRLKNIVLEHGLITSISISKSKSLQFV
ILMGEKLSIPTRDLTKLLEITAEIALTLNDFNNNTKNNKMLKUKSE
SVDSYKLSALLNNIILKQIDTIEIELKTSTVTSSELGDFSVYTKKSETIMAA
IKNLEKISTRKQELHTEENYFQSVLAMETFPITPLTVIEIQPSKKAQOLHEKSK
LVEOKLIDANVLTLHTMKQKTDISPADFTVLNQLQSTLLOLQTCVTDNLID
KKTSTNVQOOLSYIGWAEIESHQWNEFPKADVPILKILDDIKKEIQOVTOKNEE
TSLKILADVQTLLENKQSDLSPLIQHITKAGTLVGERENQKFSKNTVQKLS
SEEFKTLIDSTLLENVOLQIEISILQSNQYHQSETIKOAFDKSNTIINNLIQ
INQOYTTQPMLIAVRFLESEAKFRESENTICEITSLVSLGSLLSSTVEALKDA
LKSIDTLKEKLTAVDRPLKRELVNIRKLOKLTLLQEQDFDNKMEVDSFVPTPSR
DKVTFIQNASPKAKOYAKKALQIOAMEIDVDPEVIEDNIKANGAKNQKQISAF
QDNFSLILPDDMLSLAKETRYRSTLFTVIGPILLKVEVLESVKMKAKUKLSLL
PNGVPTPPKDFMDIHYTESNVNPHLKTINLPKSTVAHNIEHLSLSQALNSKTLPE
AVGTSLEDAHAESKCFKLEATWHDQVDRTRKIDYIEDLRNDRKHTVAQIOQS
PNRFSLEQIEINSLPKLFRDLSLENESRLASQKNEFOMLENTVKAELQYATQES
DIISNMEASLLPLAPAYTLAIPITPDKLVVENTIODKRLNTEPYOITIECLN
WLNTAKTLISLCSQKQSVQRLQDQSDINTHLNTOQFYNLEKANTDITDLSVLQNAI
STLDRKVGQKATVDSQKQKQKAMLDNKSQAQTLASLDLIMGALTALVSVAHL
STLQKADPLQKQDIESUSNTDILLRVTELIHFIFKRGFLSYEEGQKVFQRYPL
QONTROPQTEINNLRALFVLNKNDAASAWITETPLVDSNKLAVPPNKKPPLYT
CSQYKLEQLADLPDLSKVLSDNRPLAGTAQARGLDIDTVLLARAPEDTQKHAER
LTAVKNISHTONEFAMTIVCHMIKIMNDYFQNFNINVTYVNVNHTLKOIILT
MWPLIKASLCQSQSQEATSLQTLPLFLKITDLENNIYNPASHCSALLFFPQ
KWSINIQSIMEHPFSLAICKNSRARITFALAFKIIDITLILQMLTSLNPANTS
STVSLNHLNVALEFQNVSTPLEPNPSLAYAYQTQGTNIIGTKSYVQKPSPIIS
KTSALGALALFOVPVKLVDTKTPVLSPELGDMLVISELLCTQTEPTEKMEIA
PKSSLNKNQYVSPPELEFVSRQASQLHILSSSNKNNIVATIDYSTTEFNAYV
VPEKLPKQSQSCFIPKIDSLOWNNFTTFLPLVEMPSNLELHAKVEFNFTVLS
TMNVFTHLPLQEEHDSISSKSPFKIEHDINTSVNNHNNINLNNTWNSYHQY
KDVLPQDLAKLSYEPKDLQNLASTPEQIEDISELSKETDNTAKAPLYQKQK
TKKFLSPVHTKHKHTNSIIEFENTVVKVQNTCIHQGLDHDKDTNPQQISNAPCFIP
NHKVPVIFKPSQEKLVNTHTDLSPPKPKQLIANNNNIFKQSKQKHQVTOIS
KPKTIFINQDSNNPIKQPHHNPQPKLPTDQQLTSLNDIISDQTTKNLNVKPII

```

CDS

CDS

CDS

CDS

CDS
VIPNNYALNOVKLSNPSIKTKPYITLKDIQSNKTYLDESPIITILEHLDIEPI
VSISLEKRVDETKEIILEFIKTKQNIKIITNLLIHQIMKTIYLV
complement(96613. .97032)
/note="ORF 65; ECLF7; sim. to EBV BFRF3"
/codon_start=1
/db_xref="PID:g60386"
/db_xref="SWISS-PROT:Q01047"
/translation="MHLRVDPVVOGKLESDSHLVKRLIILPQNMTPEVITLM
KRNLYVFLIAQNFQDIYETIYHGIKRRKHIEGLKANLPKSAUSASLGLSTSLT
TPSLSTSTSLMPSGLSIGPSTTDTIDSKKKPKAK"
complement(97025. .98332)
/note="ORF 66; ECLF6; sim. to other HV"
/codon_start=1
/db_xref="PID:g60387"
/db_xref="SWISS-PROT:Q01046"
/translation="MDTIOERYMDNLYKFGSGCDNFKVETLNSLLQPLQIDSSDFI
KEVYGGCINEHCLPSPVFLDRCTISSELSFWCVIWDTRRTQVHKFKIKLQTC
LFRAIVVVVFPKRDLPKPKFLENWYKIINPFYKAIVTFMLNLNLSIKHPLI
QFQCLPMDLSILRKNKLCFSTLPLSPAPSORRENENLVHSDDDQSHFALMA
ALKQOQAVPGCNPLDAMIKVLCFNSMQNKYAIIPMONIEKTFDVLVKILGYNIL
SSVGLPILCKIKDKIKNTYSQHIIVCIIEGHCNLFGRGSKNLNFPPTTHVFCRD
QIKQFTICGTSGRYICSGCSQSPKFPMPVEANIRAVIANNAACMAQCAQQQDDV
VPCLMGCSICFKRTVQSLYLLYLSKIESLCCVKCGVIYNA"
complement(98245. .98952)
/note="ORF 67; ECLF5; sim. to other HV"
/codon_start=1
/db_xref="PID:g60388"
/db_xref="SWISS-PROT:Q01045"
/translation="MNSRVLVELCDIVNLYLQCPGVQIDVDRCASGPHVFTKGCTEA
ICTVKLSHLYNTEFYVFKWAHKLKSVKYPFPCFIIISNGLATLAKCFSPKRNVN
HFGHLNLDSDVLTKNSTSVILSQDDFVKFTNLVFSKDLDFHVMVFRYLIEHRQ
ALQFLVVKPRSKRVNLSLVAKTASQNFILDPFRSEETRVCIKPTLSKKNWTI
ILSLVAVAILIKRWEL"
99198. .100508
/note="ORF 68; ECRF1; sim. to other HV"
/codon_start=1
/db_xref="PID:g60389"
/db_xref="SWISS-PROT:Q01040"
/translation="MIIPQKSTLYKHRTSIECLLYNFGMPGTPETALDNLALVHTYA
ALTSTCKICQITLYSLKSNTPVAFVSEYSLCLCTCLYAPITWTSLTMAADFIEI
IKTHFTSDTFNFAQVQSLAIDQLHFYIHRCKVLSNDILSSLOFLKTTFLQG
KLGTGIPQCFCKAWIKNDCCNNTSHDLPNSLSVFCCKADQLKPLLPIIDWS
ASDLFKNVNSQEPFFYPEDIDICQCPCLLSPSLGTQKNNTISICPLCECIASHP
NAIDTQPLKTIINCENNVKLDRIISILNDELDFIQDPIKATVIONCSIOEIHK
HFFCDQCALNIKTSTNLFKIPDNLKLVLCARLATGEHLKSNLYLDCYLETAL
IFKSGTCKVGTTFLEIIRELDLLSKKHNPVTKAFOTSQIA"
100511. .101296
/note="ORF 69; ECRF2; sim. to other HV"
/codon_start=1
/db_xref="PID:g60390"
/db_xref="SWISS-PROT:Q01041"
/translation="MSSRSVKRSKSTKSRHHPPYVKLDKMFSAISSKKELGTDFLR
EMDAPICTKTILLPLDLNLSIPGRVILSPGHSSNMEFOCEKTESKNGSGDVSO
NHDLYSVTLFVYKVDVKKHAYLSLSHSMENLKSFTQPELLYIAVYVKGAGHN
VFPFPRKDCCLSLCFKCTKHIGESCLRMNDMLNPISIDYIKDVYAMFTQC
FAIORSIAEDTICESVSTLDCDELREEIVKGINALQIKDI"
complement(101533. .102417)
/note="ORF 70; ECLF4"
/codon_start=1
/product="thymidylate synthase"
/db_xref="PID:g60391"
/db_xref="SWISS-PROT:P06854"
/translation="MSTHERQHEHQVLSQVHILNYGSFKNDRTGTGTLISFGTQS
RFSLENEFPLLTTRVWRGVVLELLWFIIRGSTSKELSAAGVHIWDANGSRSLDKL
GFYDRDGLDTPVYFGWVHGAAYKGVGRDYKGEVDQLQIDITIKNTDRLML
CANNVSDIPKMWLPCHVLSQFYVDCDKLSQQLQORSADMLGVFPFNASLSLTCMI
AHVNLVGFETIHITIGDAHYVDHIDALKMQLTRTPRPFRLTFARNVSCIDDFKADD
IILEYNPHPIIKMHMAV"
complement(104685. .105189)
/note="ORF 71; ECLF3"
/codon_start=1
/db_xref="PID:g60392"
/db_xref="SWISS-PROT:Q01044"

CDS
/translation="MDLTKTVLHITDSFTEDEMYCLIFLINGICIPRSCNAVNLSDLI
ETLSKSTOWDCLMOCLVLRKIGLLNLNLFQVTKAVKOSFTQPOLETHVTLVNVN
NMLTAKDERKLCFILDQFPFRNVAAPSVILCVFSNNMLCEHVLCELCQLKCKLQIGR
SDIAKTV"
complement(105190. .105954)
/note="ORF 72; ECLF2"
/codon_start=1
/product="cyclin homologue"
/db_xref="PID:g60393"
/db_xref="SWISS-PROT:Q01043"
/translation="NADSPNLRNRAKIDSTTMKDPVNLNKLRELLLPKFTSLWFIQ
TEVVDNRITILLTWHLLCESEFELDKSVFPLSVSLDRYLCKKQTKKTLQIGRACV
LIGSKIRVVKMPTVSKLYSLSCDCTNLELINOERKDILEALKWDTEAVIATDFLPLC
NALKIPEDLWPLYEAASTICKALIQNIALLSPLGLICAGGLTTIETDNTNCRPWT
CYLEDSSILNFTSTNTRTKVDQSEAFSLYDLIL"
complement(106010. .107233)
/note="ORF 73; ECLF1"
/codon_start=1
/db_xref="PID:g60394"
/db_xref="SWISS-PROT:Q01042"
/translation="MAPRRKAKRRRHTLRSECKDKCKHVOCYVSPRRKRRKLKPOG
DDIINTHQOAAALTEQORREVEBERERBRGEGEGEGEGEGEGEGEGEGEGEAE
EKEAE
EAE
EAE
RSRLPKGSHDKVPKKYQARNKFFSOAAPSVIDLSPKSWCVVDFWGPDKLQMPQH
SLSPGAVSSGIGTPTPKGPHATGPWYFITYVTCRTQTAKEVIKAQKYEKKYPRSAK
LKASLGKSKSLPIE"
107732. .108697
/note="ORF 74; ECRF3"
/codon_start=1
/product="G protein-coupled receptor homologue"
/db_xref="PID:g60395"
/db_xref="SWISS-PROT:Q01035"
/translation="MEVKLDFSEDFSNYSNGDIYGVDPVPCVWNLISESALAF
IYVLMFLCAITGNSLVKTLFLYKRAQOSFDYLMGMFCNLNLSFLAGILMLRMLPFI
FMNTECLKLEAFNLNLSYSPFLLVFSVLCILFLICATLWVKTKLIGQVFLCCSF
VLACGLPWHVMTSYSPSSICEEDVLTQRLTKLNTFTWTFAGFLPFTVTCYS
MSCYKLFKTKSKRAEVVITMTTLLFVFCIPYIMESIDTLRVGVIEETCAKRS
ATVYGICTYMLLYCYMLPLMFAMFGLSFQRMAWCKTICH"
complement(108782. .112681)
/note="ORF 75; EILF1; sim. to ORF 03 and EBV BNRF1"
/codon_start=1
/db_xref="PID:g60396"
/db_xref="SWISS-PROT:P11282"
/translation="MAANRHGHLPLPEGLAAPTHTVYIYAKCDFSPSEECVNRPT
GRPGLTLMRORSNVEHVLTVSYGNLHQRLLTSALEQORACKTKLILLLAPQI
DYNPLTLDLHNLGQRAVIFSYGNLHQRLLTSALEQORACKTKLILLLAPQI
FTSKAVQYVVENEDDILTAALIGENNYINTETLSATRLVTDALSWAEYKPLDVTCH
AVMKPPREGVQPMCLVASPSVDYFNFKYLMITPCTAISLHMGHPYRVSQIGFHVH
SSTRPMGQCQDYLPHSSLLPNGNTAFACGYVVTSTGSGSIPVSVNEOILHAASAO
GRSLNLGAVPVVSGFLKPLPCRSSEVPNNVITHVSTGASERDNLNCRVACQFVAAY
GAYDTPSGDKSPILYRDSIDSLNRAIQKLFYOMCETPCVSSYQREFGSCSTLHL
LALVSPKMTVHISKLPEETKALRSVPVLEEDTVCAFVGYFFNCSQLFLVDQAT
VKTPSGGIIHFTDILKAGNLGAPVYILGRTNDIGICHVNDLYHPRDLVLDQAT
SSMLITVGHASVVSATLQPOBHEDESIDWMEGTSTSTISNLSHPAVASKNIIR
LDRCGNGLIAQPGIGSDAPVSDYAIICDSSMPEARLENDASQIKISKQEAORAY
AQIHKWFTKELFNTISAKYIALGEQAYKLSRNPVGVKVAIAIAINIMPGCVL
EDITLTVAAHNKQETALYRVLFACKMRELNVNLSITDASDRDPTQIDTAANT
VYVTSARVTSIERITPAKKAENALHVCLSKELTSGSVFENSTAFSSHLPLDT
KIDDMFYAVKHLISKNLVAGHDSIDGLITTAEMCFSTFVTVNLPSPALPAMY
LVSETPGALLVPEKHLSTVTTLLSERGLTWAVGTNNVNNKLSIYDNGTHLTSIN
ILNSKMWYAESEPTCEPHIESMYNDYGNMMDLKHLEDCLTKHPLQLYTCSPHPV
SVAVTFPGCDPDPVATLQAFANGFLSPITSTFELQGNLNAFSLVAGSSAESEE
GYATRIATYLLQCDLAKNLEKFFQRDLTSLCCGELGTQLLAACQVWGDTPRGRD
ISSNPETWLEPNKASHYESLWNLNHFVYPTTKSIILQALRGTTFQGLWGLVGRY
KHDAQEYIMQONGIAMSYSKSNIPYIYAMHYPRNPSSNAGVAGICSNGRSHALLY
EPALSFHTWQWQHPKPLVTPSWALMYQCMFLWCVE"

CDS
37678 a 20687 c 18269 g 36296 t
BASE COUNT
ORIGIN

Query Match 3.3%; Score 79.4; DB 17; Length 112930;

[illegible]

```

polyA_signal
5761. .5766
/Note="EDRF3 ORF"
/Complement(5793. .5798)
/Note="EDLF1 ORF"
/Complement(5797. .5399)
/Note="first methionine codon in the EDLF1 ORF"
/codon_start=1
/db_xref="PID:g330984"
/translation="MQLWLSCHWPFGRKQYKVVPLDTSSCPDRWKIEIEIAQPPGV
FVGDILQSDASARQAYLLAVQLNITDIYLRKFEASVPESSCVVQTITKRSV
RNIWNTMLSMVAGVGTIDAAKTLKLDKRAGESIAIEMLATAIVMDSDSKAWAKE
INNIISAEHERQIILVNSEVPIECETLAERKTTTTPAVSI"
6378. .8885
/Note="first methionine codon in the EDRF4 ORF"
/codon_start=1
/function="required for DNA replication"
/db_xref="PID:g330988"
/translation="MNEATEAWPKFKVLFATDGSAAEITDILGTDTNFIYVSLNH
CYIYTEVKVILICLPAPKPGGDKLEVFOLHIDTELAIPLEFVTKPLKANDLHKY
IDFKARKNFKPIDILISNPKSPKTHNSDIKSIWFAKFNLSRLKATSSSPYW
MITTGSVEPPELLAIFVFFQHNCTINTIFHLSLFEKKLGTSLIATITTEELGGV
CSTDLKATAPINFCHEIKLARDLSQAIDTSDTLRGQLMSNQDLVHYIYISFF
OCLNDIFIKYSHLNSDNIHFVPETVLAQLDENFRKDMLTYYNKSQDLYIYIHK
CIHLPDLYAGAPQDCTSFYVWAGQKNVHNLEVINTHPHNISDELNGLDLDAID
STFVNDLKDQVNESQKVPVYRCFLNKTYEIVQNDILKNVSTDVLMQENWYM
LKDEITSNISKEFTSMILRDQLKISRHEFYNRPVFNLDLHDHITSEHEI
DEIYNLCTLSLILETLOGLPVDIDTHHVFIFKSTCEKNWINDNKEKFCYCTKK
LGFRIITPLPAGVILGNSPVISFVNILNRTIKDKKLLAMYPLIMETDPPFDGIYH
KRCVRIPTYHYNSGRLERLLKFCVHPHYDNKLQIYVMSDFDINLLYHSHNPKV
KOLKAVYDIADTNENFLOKAQALPQTNHNAVERIESASHMSITDWAFAWPLFE
LIKLYLSEKVSQFVHVSFAASTGNLIKITSISGNFSCNLFKRLKTSQVRIFLSLH
TPDNCVTLTMSQCFASKNSKNCIAHMSVRYPIDK"
8881. .8886
/Note="EDRF4 ORF"
/Complement(9074. .9079)
/Note="BELF5 ORF"
<9082. .10320
/Note="5' end of coding region for EJRF1 undetermined."
/codon_start=1
/product="transactivator"
/db_xref="PID:g330986"
/translation="VLGISDDDDFSDSSSDSEEDTSPQIMKSDVTWASPPSPPEP
SPDVASVSNLKRQRSPRTWEHQSLRSVYRSPRFRGPRPRISNSSTRSKTS
WADRYREAAQRPRSPKPYSPHNGPLRNGPPRAPPLKLFDISILPKSGEPTFN
LPVSLPCOEAKTNDKYVLAQAMRHDPVSISSKQLANLPVFKPLLSVRYTPN
YYWYMRKETASNLCTAAAFDLSELCWQQQLKNDPIFSENGKDIILDSALLS
OLVHKIKMLPFCHEMOTTPQDHIVKQVCLIASNNRILDAVRYLQTSVIKSPVILL
AYAVCLPAALICTKNETOLYSHCMRLKREYRGDVMNLIHESITQHLNKCPSSTCAYT
TRAIVGTANTGTGLEFLPQ"
10425. .10430
/Note="EJRF1 ORF"
/Complement(10679. .11752)
/Note="first methionine codon in the BELF5 ORF"
/codon_start=1
/db_xref="PID:g330987"
/translation="MIQEMWTQWLVNIRASCIGALAAATPFIWCFIKSLTFTSIFSL
QISFYWALGHAVILCYITFTKEMSYIEALGACFLTMLTFTFISHLMWAPLY
TLPFVNLICLSLWPTIYDVLCPITYKVEYELGFLNALLYVYVAVNRMVLSV
IFMCPFVLFGVGFALKNFHPHFENILITCKPIFTAKNKYKGTGVNKKLVYHN
LGAVLFLVAPASVSVVQRLDIFVQWNYLFLFVLMCCMSLSPSNALCVLLET
FAVLIIVHLIDKPLSGLVIVILLMCQAIGCOIETIRTKLAGDVGPKLCL
FLCTVCNIVSVLALPCCNKSQEL"
/Complement(11147. .11152)
/Note="BELF4 ORF"
/Complement(11749. .12855)
/Note="first methionine codon in the BELF4 ORF"
/codon_start=1
/function="required for DNA replication"
/product="DNA-binding protein"
/db_xref="PID:g330988"
/translation="NADQTTHFSYGVGKPCVLOEHHKIYNIHAKVKDGLIQIEGL
TSSPSLRWSSIGELGFKINNVSEVFNCHMPENVSFNRNISPQGNFLYREL
FGCNVKTATLMSFNSRSGKPFPIKALEYCDKRVSTRTTIPAGILPVENVNSTVS
LSRVLSPKTAAMKWLREHKAHKTVMVSNSTLAVLILTVEDSTKTVLDLHIVDDMS
LCNNSDKMGVQNDQTKVNITALSALGLCKIPGVPCIKFYSEVLEVCCTPVK
LGSWDTSLQVILFCTPNLAEDTQAARVFFISPSACLTADPADTQEBTSSNQCD
SSVENSALPKRRKREKINDIKIKIAFNPLI"
/Complement(11774. .11779)
/Note="BELF2 ORF"
/Complement(11774. .11779)
/Note="BELF3 ORF"
/Complement(12966. .13883)
/standard_name="ribonucleotide reductase beta subunit"
/EC_number="1.17.4.1"
/Note="first methionine codon in the BELF3 ORF"
/codon_start=1
/product="ribonucleotide reductase small subunit"
/db_xref="PID:g330989"
/translation="MDSVKKLYTCDHVGFLTKETKNRMFPSPVQLQGDVCCVDM
LNERDLFYKFLTFGLMAEKNININIELLSQFSDSHDISHYAEQAMENIHGKYYA
NILNMLFKNITEVSYACDINMSALQELKRLWNGRVTESADKAKILLLVEGIF
FISSEFISGLFRGIMGICLANDYIARDEMLHTSAAALLYNTWTKSSERSESDIY
KLFRAVEVEFRFIAKGVSVLNVHEIRQLQATADRILLESINLNIYISLUPPENC
PLAYTSSTKSVNFFERDSDYTGTLTNDL"
/Complement(13889. .16192)
/standard_name="ribonucleotide reductase alpha subunit"
/EC_number="1.17.4.1"
/Note="first methionine codon in the BELF2 ORF"
/codon_start=1
/product="ribonucleotide reductase large subunit"
/db_xref="PID:g330990"
/translation="MSQETIISNIDMLKVSAGWDEANEISGRPLKMLKMSSTETI
SOYMSLFGLEPHILEFIQYEQEIDEVLEYRASYDFMCLRNGCIGILPAKRYDHYV
LPRTEMNGKYESIPHEFARIAACANWCMCEPLEKDTLVYVQKRWNVVEIKDMQIF
KIFYVISOVLVCATPVMRSAGVAGENUSSCFIAPTIDTEKSTISSIFGELAPLLA
SRSGVDVDTKFSFGCKNIHCLINAQVEFFNDKSVRSPVATYIEVHQIHEFL
SAKLPEPDRCSIFQGVCPVSLFFKMYESDPNGWLYLPQDAPNLRTYGLGEFEE
YLRLVSEKXKOSVTLKSLMFLINTIITKTSPIYVTSKAMNKHVWYTOGAJCSN
LCAEIVQOPKQFTSCNLANVCLPKLNSNFPYCSNTAQDESKLEYAVOAAVLI
NACILSPSTSATVQQRERSMGICHLGADVSEMGYIGLDESCLDRDIFETMYI
TAVYSSEICSVGKQGFAGFRKSLAHGVFWATWDMQPRVPMKOWHLODNIKKF
GVNFSQFIALPMTAGTQLTGYTDFYFANNKSVKSEIEMPIITFLKNVKPD
LCTVRFYGGDVMPEDEVTRYKHLTADYDCEAQMRRASIRAPVDQSQSILFLT
ENVOVASKYLKDLLGLLFRGLKTIWYCRVKTKTKLQLECLKDEHTKKQAQIVLA
DLAREPDSHKTEDACPLDQSECIACQ"
/Complement(15105. .15110)
/Note="BELF1 ORF"
/Complement(16192. .17184)
/Note="first methionine codon in the BELF1 ORF"
/codon_start=1
/product="capsid protein"
/db_xref="PID:g330991"
/translation="MPKVNSFKKIEKDVLKLLPPRPHKIFISEQKFAKSIQALS
KYTVSVNHPIDFFSNQMLDVPFRQPLYADFLQAKTIHQKEPIGSPFLTFKQBDSSS
IDIIFSPVSFYKISGLDAEFAPSVHRISFVWAGADNCIEDTVNLAELIEEGKHQFL
TPVGLPVENITSTFVTRIKSVVKGVELNNKSPENVKLLLPDLDFIDLEPNIDTLP
KHLFLYACVTYIPVLHNPVNFALFIKNSIRDLCFLKVIYSIDIIKTKYIAMQND
FYINKMSFALCKLGSSETVPRGNVHFRGSSLPTEVDFVDFADPGPMWIL"
17191. .19890
/Note="first methionine codon in the EERF1 ORF"
/codon_start=1
/db_xref="PID:g330992"
/translation="MSVKPLTELTKNLEAASNELLKTKVLMDEMLNLYSIEQLESAQN
ITDFNLNQTSQYSTTHQFLYLLKSTFTSTNYDLESIKQYMNILNVCVDTAT
TTHATSSNNLNQTVINHIOYITSNATFTGLSPBPVFNPNVIFPRCIVEEVHVCFQ
CYWHEPFOAQIPQIPNGALEKLLTQHKFLKLDYAFNLSKDAQATYLLTHEKHLVFP
LSSEYSLTLPALKQALNIYTSFTNITKSNVPLAFAKELDTATBELFLDYFI
IEALYHESYNVPQNIIEQISKTQFMIDLCNTIQIKCSNLSLTSSEIRHRELLES
CGLTECCRLQTSVLINVSFTSNWKGYTEFISLSQVLFSPFYKFLTFYHSPTS
IGHSKITELTNTVSAIESLTSKANKFSWKLAMLSFFIPKAPSKILLTYHSYSL
MKSAFSIAWAKTNVTLDTATSPHTQTOHTFKHAPVISOSEVQKVCENALQGTVEYDS
RIVNSHLFAEFETHIITPTALQNKQKRALFQLRWLIVFASDEAKGLYRRRP
LGLLQFQIEIFPHDSNAAAALNLNLDYLNIEQQLIQIYVPTIYPTIKFIELFSLKY
KQPSLETSKIQKFAETCYVKDILPFLGTNMCNTIYHIENTINIQGSPAR
LDTKALHTAIKAIQGLTESWTTISQSYKELOTATYIQTALILETEKISHQSIKVS
NPNEFIKNNTFLQCFKCKYNTIANLITNSHFNLTFRYQIFPELIPITVTKILNFN

```


Query Match	3.3%	Score 79.4	DB 17	Length 3720
Best Local Similarity	47.8%	Pred. No. 7.6e-05		
Matches 230	Conservative 0	Mismatches 251	Indels 0	Gaps 0
QY	362	ATGAATTCGTTACACGCACACATCTTTGAGGAGAAAGACTTGCACACAGAGAGATGGACCCA	421	
Db	1965	AGAAGGAGAAGGAGAAGAAAGCTGAAGAAGAAAGCTGAAGAAAGAAAGCTGAAG	1906	
QY	422	CAGGAGAACACACAACAGGAGGATGATGAGTTCTTATGGCGACGTGATGTAGATCATAGAT	481	
Db	1905	AGAAGAAGCTGAAGAAGCTGAAGAAGAGCTGAAGAAGAAAGCTGAAGAAGCAAG	1846	
QY	482	TTGAGACCTTGGAAACCTGAAGTATCTCATGAAGAAACCGAGCATAGTTACCACGTGGAA	541	
Db	1845	CTGAAGAAGAAGAGCTGAAGAAGAAGCTGAAGAAGAAGAGCTGAAGAAGCTGAAG	1786	
QY	542	AGACAGTITTCACAGACTGTAATCAGATATGGAAGAGAGATGATCTTGACGAGGAAATC	601	
Db	1785	AGAAGAAGCTTGAAGAAGCTGAAGAAGAAGCTGAAGAAGAAGAGCTGAAGAAGAGCTG	1726	
QY	602	CAGATTCCAGTGAACCACTAGTACAAGATGAACATTGCCACCATGATACAGATGATCTAA	661	
Db	1725	AGAAGAAGCTGAAGAAGCTGAAGAAGCTGAAGAAGCTGAAGAAGAGAGCTGAAGAAG	1666	
QY	662	CATACCAAGTCTATGAGGAACAACGACGTATATGAACCTCTAGAAAAATGAAGGATAGAA	721	
Db	1665	CTGAAGAAGCTGAAGAAGCTGAAGAAGCTGAAGAAGAGCTGAAGAAGCTGAAGAAG	1606	
QY	722	TCACAGAAGTAACCTGCTCCCTCCCTGAGGATAATCTCTGTAGAAGATTACAGGTAATCTG	781	

Search completed: May 14, 1999, 01:44:28
Job time: 1688 sec